DOCUMENT RESUME

ED 254 592 UD 024 108

AUTHOR House, Ernest R.; Husen, Torsten
TITLE Reviews of Reports from the National Institute of Education on the Compensatory Education Study (6 Volumes).
INSTITUTION National Academy of Education, Washington, D.C.
PUB DATE 79
NOTE 131p.; Proceedings of the National Academy of Education Vol. 6, p. 358-486, 1979.
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS *Compensatory Education; Cultural Context; Disadvantaged Youth; *Educational Philosophy; Elementary Secondary Education; Equal Education; Evaluation Criteria; Federal Programs; *Program Administration; *Program Effectiveness; *Program Evaluation; Research Problems; Student Development
IDENTIFIERS Elementary Secondary Education Act Title I; National Institute of Education

ABSTRACT

Two reviews of NIE's study of compensatory education are provided in this document. The conclusions of the first reviewer (House) are that: (1) the study is valid because it meets the standards for the evaluation of a public program; and (2) the compensatory programs, especially Title I, are very good. Other conclusions are that the funds' allocation requirements are necessary but the program development requirements are excessively regulated, program management is efficient but monitoring is confused, and how much students develop in these programs remains hazy. The second review, "Evaluating Compensatory Education" (Husen), takes a broader perspective, based in part on the reviewer's earlier participation in the Organization for Economic Cooperation and Development's review of U.S. educational policies. Against this background, the reviewer discusses the role of education in the United States, the philosophy and strategies behind compensatory education, Title I targeting, and evaluation problems. This review concludes that: (1) in looking at the outcome of intervention programs, a broader perspective is needed than appeared in some of the NIE evaluation efforts; (2) well-planned small-scale longitudinal studies conducted where different models of compensatory education are operated would prove more than large-scale surveys; and (3) compensatory education programs are only small changes in the larger context of formal education, and minor changes in input should not be expected to produce dramatic results. (CMG)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

The National Academy of Education

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

B. Bennett
Nat. Academy of Education

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Proceedings of the
National Academy of Education
Vol. 5, 1978
Copyright © by the National Academy of Education
All rights reserved.

Ernest R. House
University of Illinois**

Compensatory education has been the largest and most controversial area of federal involvement in education since its beginning nearly two decades ago. The evaluation of compensatory education has been no less controversial. Both touch deep-seated value and interest conflicts in society.

In 1974, the Congress instructed the National Institute of Education (NIE) to conduct a study of compensatory education, especially those programs funded by Title I of the Elementary and Secondary Education Act, which has provided over $2 billion annually for educational programs for low-achieving students in school districts with low-income families. The NIE mandate grew out of Congress's frustration with past evaluation efforts.

**Ernest House wishes to acknowledge the helpful comments provided by Gene Glass, Paul Hill, Douglas Weir, Gordon Hoke, and staff from the Illinois Board of Education -- Tom Kerins, Connie Wise, Rose Maye, and Jack Fyans.

By 1974, the evaluation of compensatory education in general and Title I in particular had indeed been inadequate. After several years of federal programs, and several hundred millions of dollars spent on the evaluations, Congress had little information on which to base decisions. As a member of the staff of the House Education and Labor Committee put it: "Incredibly enough, both the sponsors and opponents of that provision found that almost no data existed on how much money from the Title I allocation was then being spent on basic skills" (Cross, 1979, p. 16).

The reasons for this inadequacy are found in the history of the Title I evaluation and in the type of evaluation that has been attempted for the past decade. Put simply, federal officials in the Office of Education have attempted to evaluate these programs by an inappropriate approach to evaluation and when the approach fails, they modify the evaluation slightly and attempt the same approach again, and it fails again. This has occurred repeatedly over the past fifteen years (McLaughlin, 1975). The result has been either no information at all or evaluation data that are severely criticized as invalid by evaluation experts. Controversy succeeds controversy.

Frustrated by this problem as early as 1974, Congress tried two solutions. One was to write into the Title I legislation itself a requirment that the Office of Education develop "models of evaluation" and that local school districts evaluate their programs using these models. As in many cases where Congress tries to legislate the detailed conduct of technical activities, the legislation led to questionable results.

The Office of Education developed three models of evaluation which emulated the unsuccessful approach to evaluation that the Office had been using. Although the legislation was enacted in 1974, there is still little usable information on which to base decisions. In addition, the evaluation models themselves are under attack by evaluation and measurement experts, as well as by the state evaluators whom the models are supposed to help (Linn, 1979; Jaeger, 1979; Wiley, 1979). The state evaluation directors have strongly protested the imposition of these particular models, yet the work goes forward. In 1979, use of the three models was mandated for all Title I programs.

The other Congressional solution to the problem was to instruct the National Institute of Education, an agency separate from the Office of Education, to conduct a study of compensatory education, especially the Title I programs. It

is my happy duty to report that the federal government has redeemed its honor and its competence in this study. It has conducted an evaluation that was helpful to the Congress, that was valid, that was worthy of the name.

The former staff member of the House Education and Labor Committee commented:

> More recently the attention of Congress with respect to the reenactment of Title I has been focused to a remarkably large degree by information generated by the study done by the National Institute of Education. . .The NIE study has been remarkable in terms of the quality and quantity of information that it has produced . . . One of the more specific products to emerge from that study was a substantial revision and reorganization of the Title I statute. . .At this point one could say that the NIE study has been the single most productive piece of evaluation work ever undertaken on a federal education program. It may well be one of the most productive evaluation works of any executive branch agency. (Cross, 1979, pp. 20-25.)

This burst of enthusiasm is understandable after so many years of failure. The intention of this review is to analyze the NIE study, to

judge its worth, and to assess its strengths and weaknesses. I shall try to show why it is a good study and to outline the critical decisions made by the evaluators and officials that kept it from going astray. I suggest in some places how it might have been made even better than it was.

Judging Evaluations

What is it that makes an evaluation good or bad, valid or invalid? Suppose the evaluator and the decision maker are the same person, as when a person is deciding to purchase a car. What is the minimum standard for the evaluation? In this case one would say that the evaluation must be true. If the evaluation were not true, even though conducted by an individual for himself or herself, one would say the evaluation was invalid. There are many different ways in which individuals determine truth, and, indeed, these different methods distinguish evaluation approaches. But, whatever the method, an untrue evaluation would be considered invalid by everyone. It would make little sense to say, "This is a valid evaluation, but it happens to be untrue." (Validity I take in the sense of being worthy of recognition.)

Now consider the case when the evaluator and decision maker are separate persons. Perhaps the decision maker consults Consumer Re-

ports to make a decision. It is not enough that the evaluation be true; it must also be credible to the decision maker. An incredible evaluation is likewise invalid, i.e., not worthy of recognition. Again, there are different ways in which credibility is assessed. Again, it would not make much sense to say, "This is a valid evaluation but it is not credible."

Finally, consider a third case. In this case, the evaluator is not evaluating something for an individual's private use, but is evaluating a public program, a social program. In this case the evaluation must not only be true and credible, it must be normatively correct. This is the most stringent demand of all, for one has passed from the private into the public realm, and there are obligations to the public at large and not simply to private individuals. This is the area of least agreement. I will not make an extended argument here but will simply contend that an evaluation of a public program must be democratic, fair, and impartial. If it is undemocratic or unfair, it is invalid. It makes little sense to say of an evaluation of a public program that it is valid but that it is unjust, unfair, or partial. (For development of these ideas, see House, 1980.)

Why this is so is because of the role of evaluation in our social decision procedures, in determining who gets what. To be valid a pri-

vate evaluation must be true, an interpersonal evaluation must be true and credible, and a public evaluation must be true, credible, and right, i.e., democratic, fair, impartial, and so on. The burden of those who presume to evaluate social programs in the public interest is not a light one -- nor should it be. It seems only fair and just that evaluators should meet standards that they expect of others. I shall attempt to show in this review that the NIE study meets these standards of validity.

In fact, I would contend that the success, acceptance, and utility of the study is significantly dependent on its meeting these validity claims. In the words of a Congressional staffer:

> . . . Congress is generally not concerned with the technicalities in the development of evaluations. What the Congress wants is information on which to base policy and political judgments about programs. The primary concern of Congress is that the information that is produced can withstand a professional critique and that the data will produce findings that are at least somewhat conclusive. (Cross, 1979, p. 19.)

Although part of the judgment as to whether the evaluation is <u>good</u> depends on whether it can be used by its intended audience (in this case,

the Congress), utility of the information alone is not a sufficient criterion to determine the validity of the evaluation. Congress may make good use of information which is false. Or an evaluation that is good for Congress may not be good for the country as a whole. The Congressional interest may not be identical with the public interest. An evaluation could be good for Congress's own purposes and interest but invalid in that it did not serve the public interest. Thus utility for an intended audience, while important, is an insufficient criterion when evaluating public programs, even when the audience is the government.

From the quote above, it would appear that Congress wants relevant information and also information the truth and credibility of which are assured by professional consensus. Another criterion -- that of conclusiveness -- I will not address other than to say that government officials often expect a conclusiveness that is beyond the power of formal evaluation or social science to deliver. (For a discussion, see Lindblom and Cohen, 1979). Congress is actually looking for practical information that can be used to improve programs. The NIE study, while inconclusive in many ways, was able to provide such practical information and hence satisfied the Congress. More elusive are enduring generalizations about social programs and social phenomena. Pursuit of these has led to incon-

clusive evaluations and to Congressional frustrations.

In summary, the conclusion of my review is that the NIE compensatory education study is valid because it meets the standards for the evaluation of a public program, and it is a good evaluation because it was used by and satisfied its intended audience, the Congress. Its value to Congress was partially dependent on its validity. It would have been a valid evaluation even if, through some strange circumstance, Congress had never seen it or used it, in which case it would have been of no value to them.

The Charge

Section 821 of the Education Amendments of 1974 instructed the NIE to undertake a "thorough evaluation and study" of compensatory education programs. Such study had to include the following:

1. An examination of purposes and effectiveness of these programs;

2. An analysis of means to identify children who have the greatest need for such programs;

3. An analysis of the effectiveness of

methods and procedures for meeting the educational needs of children, including the use of individualized plans and teacher training programs;

4. An exploration of alternative methods of distributing funds in an equitable and efficient manner that ensures the funds reach the areas of greatest need;

5. Establishment of not more than 20 experimental programs to carry out the above purposes.

In addition to allowing four years for completion, the law instructed NIE not to report the results to any other group, e.g. the President or the Commissioner of Education, before it reported to the Congress. Specific Congressional concerns in passage of the act were the effects of altering the definition of poverty, of using test scores to distribute funds, of requiring individualized plans for all students in Title I programs, and of requiring school districts to spend 85% of their Title I grants on basic skills instruction.

From these concerens NIE designed a wide focus study, one purpose of which was to assess the past effectiveness of the program and the other to examine specific alternative proce-

dures. So the study was both an evaluation and a policy study. The wide focus was the most critical decision -- and a wise one. Past evaluations had foundered by using standardized achievement scores as the exclusive indicators of program outcomes. In both correlational and experimental designs, the Title I programs repeatedly failed to show large increases on standardized tests. Measurement and evaluation experts criticized the narrow focus of the outcomes, the experimental designs, the instruments employed, and the statistical techniques, thus throwing the evaluations into question.

By contrast the NIE officials concluded that compensatory education had a number of important goals, most of which were not captured by outcomes on standardized achievement tests, which were themselves considered by many to be invalid measures. Although there were many motives for originating compensatory education programs, such as eliminating poverty, the specific purposes clearly discernible in the status and Congressional statements were to provide financial assistance to districts with low income children, to fund special services for low-achieving children in the poorest schools, and to contribute to the overall development of students. Members of Congress themselves differed on the priority of the three fundamental purposes. There was no reason to assume (as past evaluators had done) that the programs existed

solely to enhance academic achievement as measured by tests.

Furthermore, federal involvement operated through state and local agencies which had their own goals. The evaluation had to consider implementation as a factor. This insight into compensatory education programs, particularly Title I programs, as having multiple and possibly conflicting goals and as being implemented by other agencies with their own structure and purposes, is what distinguishes the success of this evaluation from past failures. It is essentially a political insight, and it is significant that the director of the study (Paul Hill)[1] had a background in political science rather than economics, engineering, accounting, psychology, or even sociology. The temptation in all these disciplines would be to secure overly reductionist outcome measures. (Of course, political science is also reductionist in its focus on power and authority.)

The NIE strategy was designed to produce a "complete understanding" of how Title I operated by focusing on funds allocation, service delivery, student development, and program administration, and by employing criteria derived from the fundamental purposes. The study did not provide a complete understanding, but it did provide information that Congress could and would use to judge and improve Title 1. The in-

itial definition of the problem derived from the Congressional charge was critical to the success of the study. Other, less fortunate, interpretations were possible. Although most evaluation discussions focus on technicalities of design and statistics, it is in the initial conceptualizing of the problem that the essential features are determined, usually beyond remedy.

Needs

A reading of the original Congressional mandate reveals how easily the NIE study could have gone astray. The mandate is implicitly written from the "needs assessment" paradigm: determine the purpose of the programs, analyze means of identifying children in need of such programs, and analyze the effectiveness of methods for meeting the needs. Needs is a complex and unquestioned concept, and when used as in "children's needs," a clear moral mandate. The existence of a need is used to justify acting in a certain fashion. Most often, need is used to indicate the discrepancy between some target state and the actual state of affairs. (See Scriven and Roth, 1977, and Roth, 1977, for a half-dozen different ways in which needs are defined.)

It is clear enough that children need vitamin C or so many calories a day, but the fur-

ther one moves away from biology, the less certain the concept of need becomes. For example, the child needs a strong self-concept (mental health), but people disagree immediately as to what this entails. Typically, needs assessments in compensatory education programs are conceived as the need to reach the average level of educational achievement, as operationally defined by a standardized achievement test.

Such a construal of needs could have resulted in another evaluation such as occurred in the past. The study might have been designed as a large-scale field experiment to see whether Title I programs improved children's test scores or brought them up to grade level on standardized tests. An experimental design would be necessary for determining whether the expected gain was attributable to the programs and not to other causes. Yet there are many reasons why children do not perform at grade level on tests, including family background, past history of disadvantagement, and even the nature of the tests themselves. This type of evaluation makes the federal program solely responsible for elevating test scores when there are many causes of low test scores. The programs are likely to look like failures. They are, after all, only supplements to the children's education.

Thus the net effect might be to make the programs and those who run them appear incompe-

tent. President Nixon justified no increase in funding for compensatory programs because he said educators did not know what worked, a direct reference to past evaluations. Often the children themselves were blamed for the failure. Some researchers have concluded that certain children are intrinsically incompetent because compensatory programs have not succeeded, according to this type of evaluation.

In this case the Congressional mandate instructed the researchers to examine the purposes of the programs and to analyze the means for identifying the children who have the greatest need for the program. The researchers defined the set of fundamental purposes as providing financial assistance, funding special services for low-achieving students in poor schools, and contributing to the cognitive, emotional, social, or physical development of these students. Needs were not reduced to test score performance, although there was a testing component to the study. The financial assistance to poor students and the provision of special services were conceived as positive outcomes. In other words, the needs were for more money for the education of poor students and for special funding for them, as well as a need to enhance their overall development.

Other possible purposes for the programs would have been to eliminate poverty and to re-

distribute wealth. An egalitarian (and I am one) might hope for the programs to reduce poverty among the students. But, again, since poverty has many causes, many of which are external to the school's influence, expecting the schools to alleviate poverty puts an extraordinarily heavy burden on the programs and one the programs probably cannot carry. It would be interesting to know the degree to which the programs alleviated poverty, although how this might have been determined in a four-year evaluation of public school children is not at all clear. What the study did do is examine the immediate redistributive effects of the programs.

One of the internal conflicts in compensatory programs is that they are supposed to compensate for both poverty and educational disadvantagement. These two goals do not always coincide, and many decisions in the study reflect a weighing of the two, and a trade-off of one against the other. This balancing of criteria applies to the evaluation of any complex social program.

In summary, what one takes as the purposes of the programs will very much influence what one takes as the children's needs and will alter the shape and results of the evaluation. Children's needs could be construed variously as the need not to be poor, to receive extra educational help, or to perform at grade level on a test.

(There are hidden, and perhaps implausible, presumptions that the higher test score will lead to better things in life later on.) In this study I believe that the researchers have defined the problem in such a way that the results were useful.

Funds Allocation

One of the three fundamental purposes of compensatory education, as defined by the NIE Study, was to provide financial assistance to schools in relation to their numbers of low-income children. The NIE evaluators conducted an in-depth analysis of the distribution of the Title I funds. Did the funds go to districts and schools with poor children? The persistent pursuit of this question, I think, makes the NIE study democratic.

Democracy means that everyone's interests are considered, and in some cases justice is defined as attention to the interests of those who are most disadvantaged. In this case the primary interests are those of the children who are poor. Whether they, in fact, received the funds that were directed to them was a critical question. If it is an obvious one, it should be noted that up to the time of the NIE study, little information existed as to the destinations of the funds. One may also wonder whether the

funds do any good. First, though, they must arrive.

A just compensatory education program, justice being democratically conceived, will deliver the funds to the disadvantaged effectively, and a just evaluation will investigate whether the program does so. Poverty programs that help the upper classes more than the lower are all too common. Perhaps the greatest strength of the NIE study was its concern about these issues.

Who gets the money? This question is not as easily answered as one might think. According to the NIE study, the funding pattern of Title I is complex. By law, school districts are supposed to receive 40 percent of their average expense for each child between the ages of 5 and 17 identified as eligible.[2] In fact, districts receive only about 17 percent. Federal rhetoric is stronger than federal action.

The complexities of governmental jurisdictions complicate the allocation process. There are three steps. First the federal government distributes monies based on county census data. The states in turn distribute to school districts within counties, and the separate districts distribute to individual school within their jurisdictions. Eligible children are identified as those below the poverty level,

those from families receiving dependent children aid, and those supported by public funds in foster homes and institutions. Poverty level is defined by the Orshansky index, which estimates the cost of food for an adequate diet, then multiplies by three, one-third being the proportion of food costs in the average budget.

The distribution effects of the Title I funding formula were that counties with the most eligible children were indeed receiving the most money. There is a clause in the formula that bases funding partially on average costs of educating so that central cities and suburbs in the Northeast were receiving the most dollars per eligible child while rural Southern counties received fewest dollars per child. This adjustment seems reasonable since educational costs do differ, and there is no cost of living adjustment built into the poverty definitions.

The redistribution effects of Title I can be seen by the fact that when <u>all</u> school children are included (as opposed to those eligible) the central cities receive $43 per child, the suburbs $20 per child, and the nonmetropolitan areas $40 per child. In other words, the suburbs receive less because they have fewer eligible children. Similarly, the Southeast region is the most favored region of the country, according to the study. As to race, although most elibible children are white, only 25 per-

cent of the white children are from low-income families, compared to 42 percent of black children and 28 percent of Spanish-surnamed children. Overall, half the funds that would have gone to suburbs on the basis of total school population go to central cities and nonmetropolitan areas. Central cities also gain 10 percent more from the cost component than they would by strict child count. There is a limit put on the cost factor that stipulates that no state may receive less than 80 percent or more than 120 percent of the national average per pupil expenditures. This cost limitation favors the South and penalizes the Northeast.

The other factor that influences the amount of funding is the number of children in families receiving state aid and the number in institutions and foster homes. Most of these children are in the five large industrial states. This factor results in a transfer from nonmetropolitan areas to cities. The South is the big regional loser and the Northeast the big gainer. Thus, different factors in the overall formula have quite different effects.

Allocation to counties is only the first step. The states must allocate funds to districts within counties. Most states follow procedures dependent on the federal formula. However, five states allocate funds entirely on the basis of dependent children. This results in

shifting an extra 10 to 16 percent of the funds to cities which contain high percentages of minorities. These allocations are based on data that are several years old. None of the allocation procedures yield totally consistent or current information.

Within a school district, funds must be allocated to schools in which there are high concentrations of children from low-income families. There are a number of ways by which districts allocate funds to schools, and districts are given great latitude in which procedures they may follow. Eligibility can vary considerably, depending on which indicator of low income is used at the school level. Districts can increase the number of eligible students by using different measures for different schools. The NIE researchers seem to disapprove of this inconsistency, but I think it merely emphasizes that measures on which any funding is based are at best only partial indicators of what they supposedly represent. Obviously, consistency achieved by using only one measure of low income will be purchased by excluding some children identified by the other measures. In fact, I would encourage the use of multiple measures for each school.

Once schools are identified, almost all of them receive services. In half the districts the amount of funding is determined in propor-

tion to the number of low achievers, not on the basis of low income. In many districts the basis for allocation decisions is not clear. Educational "need" is cited by some.

The idea of serving both children from low-income families and low achievers is built into the program itself. Funds awarded to a school because it is in a low-income area may go to helping low achievers in the school who are not from low-income families. Personally, I see no reason for limiting the services to low achievers. It seems to me that a gifted youngster who is from a low-income family deserves extra help, too. So certainly should low achievers, whatever their income level. It would make more sense to target the funds to all children in low-income schools, since all the children in that school have a disadvantage, whatever their achievement. One could have separate programs for low achievers. The current system penalizes some poor children. There is no reason to believe that poverty has not stunted their life chances, even though they score above some arbitrarily defined test level.

Various suggestions have been made for changing the definitions of poverty. Most would result in significant changes in allocation. Census surveys tend to underestimate poverty, and more recent surveys have indicated that the South has relatively less poverty and the North-

east relatively more than before. The groups studying this issue have reached no recommendations. The measurement of poverty itself (like all measurement) presents complex problems.

Title I funds are supposed to supplement rather than supplant local funding. In other words local districts are not supposed to replace their spending with federal funds. By comparing spending patterns in similar districts with and without Title I funds, the investigators estimated that a Title I dollar increases local spending by 72 cents. This compares favorably with 41 cents on the dollar for other federal programs, 25-40 cents for revenue-sharing funds, and 13 cents for state aid. In other words, Title I funding is comparatively quite effective in raising spending on education, rather than merely providing tax relief.

Other NIE analyses assessed the redistributive effects, the degree to which Title I provided more money to low-income districts. Title I was markedly more redistributive than other state and federal programs. It allocated 5-1/2 times as much aid per pupil to districts with the highest poverty rates (grouped by quarters) than to those with the lowest. The ratio for other federal programs was 3-1/2. Title I may be the most redistributive federal program. Title I funding has a correlation of -0.76 with county per capita income. By contrast, state

aid programs were not very redistributive, showing a ratio of only 1.2 within-state.

On the other hand, Title I does not equalize expenditures. Low-spending districts are not necessarily the low-income districts. Equalization programs provide more money to low-spending districts. The cost factor in the Title I funding formula gives more money to high expenditure districts. Title I has a modest equalizing ratio of 1.6 of aid per pupil of the lowest to the highest quarter. A strong equalizing formula would likely decrease the redistribution effects. All in all, Title I is strongly redistributive in effect, aiding those districts not favored by state and local revenues. In terms of social justice, the Title I programs are commendable, and the NIE study is to be commended for investigating these issues.

Allocation by Testing[3]

A particular stimulus for the NIE study within Congress was that Representative Albert Quie (R-Minn) was interested in using test scores rather than poverty indices to allocate funds. Some people have argued that achievement is the "ultimate aim" of compensatory education and that test scores are a better basis for distribution. Others have argued that a fundamental purpose is to channel funds to the poor

rather than to low achievers.

NIE was instructed to investigate this issue and carried out a series of substudies that were particularly well conceived. The investigators reasoned that such a testing system should be technically accurate and not susceptible to manipulation. They recognized that any setting of a minimum skill level is arbitrary and that there are no technical procedures for translating standards into test items. Furthermore, the test cut-off levels have strong effects on the distribution of funds.

The investigators concluded that tests would have to be administered at least every 4 years since test scores fluctuate considerably. To prevent various forms of cheating, the tests would have to be administered by trained proctors under strictly monitored conditions. No current testing system is adequate to distribute funds among states, and few state testing programs are adequate to allocate among districts within states. The National Assessment of Education Progress was designed to produce national rather than state-by-state estimates and does not test in all states. Current testing programs are adequate to the task of allocating to schools within most districts.

Hence, allocating Title I funds by test scores would require the establishment of a new

testing program. Assuming that the testing was limited to reading, to 9-year olds, and to 10 percent precision, a 3-year testing program would cost $7 million for one set of results. For each grade level added, costs would increase by 20 percent and by 50 percent for each additional school subject. A testing program of 5 percent precision would cost $16.5 million, according to the NIE study. (The sampling error of the census data on which the poverty indices are based is very small.) To produce estimates for each school district in the country would cost $53 million. I would not argue with those figures other than to say they are conservative estimates made several years ago before the huge increase in travel costs. Half the total costs are in the data collection and would be substantially higher today. In any case, the maintenance of such a testing program would be costly.

Assuming a cut-off at the fifteenth percentile, a figure which would yield the same number of students as current Title I funding procedures, NIE investigated the effects of switching from current allocation procedures to test score allocations. Twenty-three states would have increases or decreases of eligible students in excess of 15 percent. Many of the changes would be quite dramatic. Generally, nonmetropolitan areas would lose, and central cities with high black populations would gain. The higher the cut-off score, the more the suburbs would gain.

If the cut-off were raised above 15 percent, the distribution of eligible children would resemble the distribution of the general child population. There would be no significant changes from region to region.

Distribution of funds within states would require uniform regulations for all states, since state testing programs differ significantly, and it would necessitate major changes in those state programs, according to the NIE study. My own judgment is that the rules and regulations would be far more cumbersome and directive than expected.

To determine what effect such an allocation might have within school districts, 13 school districts were given dispensation to allocate funds internally by test score. The overall effect was that they served more schools and students than before but did so less intensively. They concentrated less on poor children and on minorities. The teaching time and strategies were the same. As the number of students served by the programs increased, students were included from higher income areas. On average, there was a 71 percent increase in the number of students and only a 44 percent increase in the number of teachers. Students spent less time in special instruction. The average time loss for each student was 14 percent, due to a 13 percent reduction in the staff/student ratio. The study

of those demonstration districts lasted two years. NIE investigators concluded that one might reasonably expect the intensity of instruction to decrease still further since many of the activities were supported with extra funds not available over a longer period of time.

Although NIE took no stand on using test scores to allocate funds, I find their studies a strong negative assessment of such a prospect. To allocate so would require the establishment of another large and expensive national testing program. It would require uniform regulations and substantial changes in current state testing programs, plus a bureaucracy to enforce such changes. Even then, the data would be based on only a thin slice of achievement, perhaps one subject at one grade level. Funds would be shifted to students from higher income families and away from the poor. The higher the cut-off level moved -- certainly a strong pressure -- the more the funds would resemble general aid.

Within school districts such an allocation would also shift services to the wealthier student and would result in less intensive services for the poor. All things considered, there are remarkably few reasons for using test scores to allocate funds. Compensatory education would cease to be a program for helping the poor. There may be good reason to have programs for

low achievers regardless of income, such as for handicapped students, but there is no reason to purchase these at the expense of the poor children.

For the most part, I think these studies (policy studies rather than evaluations) examining the effects of a shift to test score allocation, are quite well done. They explore the relevant issues in a clever way. Policy issues are elusive since they anticipate future effects. These studies are conclusive, convincing, and succinct. Ordinarily, when one does a feasibility study for employing a new method, the study is highly likely to favor the new method.[4] That was not the case here. The investigators displayed an evenhandedness in dealing with a critical and politically sensitive issue.

Administration[5]

Another often neglected aspect of a program being evaluated is its administration. The administration of a program cannot only account for its efficiency, but can affect its justice and fairness. The cornerstone of justice, after all, is consistency in application of principles and rules. One cannot have a just program where regulations are wrongly or inconsistently administered.

The NIE study provided for a thorough examination of the Title I legal framework and its administration. At the federal level there are two types of requirement, those for funds allocation and those for program development. NIE contracted with the Lawyers' Committee for Civil Rights Under Law to do a survey of the legal requirements. The criteria for this assessment of the legal framework were whether the regulations were necessary to accomplish the program purposes, whether the regulations were internally consistent and consistent with the statutes, whether they were clear and comprehensible, and whether they were unduly restrictive of local district flexibility -- an excellent set of criteria.

The funds allocation requirements included seven rules, such as targeting of schools, eligibility of students, and funds being used to supplement instead of supplant. The investigators concluded that funds allocation requirements were necessary because of strong pressure on states and local districts to use the funds in other, noncategorical ways. I would agree with this judgment. Without strong stipulations the funds would almost certainly be used for general aid rather than to help the poor.

The program development requirements were a different matter. They included needs assessments, objectives, evaluation, and parent in-

volvement. While these things may be desirable, they are certainly not necessary to accomplish the purposes of the program, according to the study. In fact, one may meet these program development requirements and still have a low quality program. There is little evidence that any of the designated requirements will actually result in a better program. I would judge that some work against it. There is an unfortunate tendency on the part of the government to specify not only that something be done, such as providing programs for poor children, but also to tell people exactly how to do it. As often as not, the advice is misguided and ill-conceived.

There is another difference too between the funds allocation and program development requirements. With funds allocation there are strong incentives on the part of the state and local agencies to use the funds in other ways than those intended. Hence, the restrictions make some sense. However, there are no incentives to deliver inferior quality programs. The program development requirements are based on the supposition that the federal government knows something that the local agencies don't. Although some of the requirements are good ones, prescribing the contents of local programs is a case of over-regulation.

As to consistency and clarity, the NIE investigators found the regulations generally con-

sistent, although the role of the parent advisory councils and that of the local and state governments in monitoring could be made more specific. On the other hand, the operational standards were judged not to be clear. Clarity was defined operationally as a Title I administrator's being able to comprehend the full range of acceptable practices from the regulations. State administrators had to invest an inordinate amount of time to understand, especially regarding the requirements of needs assessment, evaluation, and parent involvement.

This lack of clarity resulted in overly restrictive interpretations by the states. Because of fear of violating the requirements, some states wrongly concluded that they had to have pull-out procedures or that they could not approve the hiring of nurses, for example. The lack of clarity in regulations actually led to less flexibility for local districts and to promulgation of inferior programs, a point to be made later about pull-out procedures.

At the federal level itself, the investigators found that monitoring and enforcement regulations of the supplement requirement were neither clear nor consistent. Two different groups inside the Office of Education charged with the monitoring of the supplanting regulations were in conflict over the proper criteria for supplanting violations. As a consequence,

the monitoring of this important regulation had ceased. States did not know whether or not they were in violation. One can see that this kind of information is valuable for an evaluation to uncover because something can be done about it immediately.

Overall, the federal management of Title I absorbed $5 million and employed 100 people. The states spent $20 million for management and the 14,000 local districts, $120 million. The state agencies received one percent of the state funds for administration, and the local districts spent about four percent of their funds for management. These are not bad figures for such a large program. State staffs averaged 18 full-time people, although this varied considerably form 3 to 101 persons. Generally, the states with poor management performance, as judged by such things as audit exceptions, had small staffs. The successful state managements spent much time with the staff of local programs.

NIE was also instructed to investigate the state compensatory education programs other than Title I to see if there were other models of good practice the federal government might adopt. This investigation seemed rather cursory to me. Fourteen state compensatory programs provided a sum of money equivalent to 20 percent of the Title I funds, but 68 percent of this was in New York and California. According to the

study, the services provided by these state programs were very similar to those of Title I. Even the administration of these programs was conducted from Title I offices. In the opinions of the state directors of these programs, Title I was essential to the success of the state programs. Few states would have established their own programs without the federal example, and few state agencies could resist pressures to discontinue. Ironically, some of the strongest fiscal pressures on the compensatory programs (which ruled out increased funding) were federal incentives and requirements for handicapped children.

Some effort, though not a lot, was put into studying the administration of local programs. At the district level, planning was done in a few meetings in the spring. Few teachers or principals attended these meetings, and teachers did not feel constrained by plans developed there. Presumably, these activities were to meet the program development requirements. Plans did not relate to the classroom level. Apparently the real decisions were made at the school level meetings, which involved the participating principals and about 90 percent of the teachers. In these meetings instructional techniques were discussed. But information on this process is sparse and fragmentary.

The elevations required of the programs

were not disseminated below the district level, and when they were, teachers felt the standardized test scores provided little information of relevance to the classroom. Fully 95 percent of the evaluations relied on such tests, and, in fact, the use of such tests is now mandated by the Office of Education. Evaluation of compensatory programs continues to be frequently dysfunctional. It is unfortunate that useful evaluations cannot be done. Again, the regulations specify not only that an evaluation be done, but how it should be done, usually counter to professional advice.

Another program development requirement and problem area was the Parent Advisory Council. The regulations provided that the majority of the members must be parents of participating children and be selected by parents from that area. The idea was to give poor parents some say in the education of their children. This was easier said than done. Most councils met less than once a month, and they were purely advisory 90 percent of the time. In half of them, the principals dominated the choices to be made, according to the study. In spite of this, 71 percent of the members reported being satisfied with their role. Some did perceive some unfairness. Again, the NIE account is thin.

Overall, the evaluation of the management of the compensatory program is quite strong,

well-considered, and unusual in its focus, especially the part dealing with federal regulations. This should be an important component of similar evaluations. By addressing the concerns of those who must live by these regulations, it substantially increased the utility and justice of the study. The major weakness was a lack of similar thoroughness at the state and local levels.

Compensatory Education Services

The NIE study was actually a series of separate studies, more than 35 altogether, which were organized and published in 7 separate volumes. Each substudy addressed a separate set of questions. The logic of this design is another reason for the success of the overall study. An alternative way of proceeding might have been to design a massive experiment, the results of which would have left unanswered most key questions. In other words, a whole set of relevant questions was explored by the NIE study, thus sustaining and eliminating various hypotheses about compensatory programs. Social issues can rarely be resolved by a single massive experiment. There are too many separate issues to address by a single approach.

The first volume of the NIE study focused on a description of the compensatory programs,

especially Title I. This was necessary because no current information representative of the nation existed on the selection of students, the educational services, or the costs. The multitiered reporting system in which local districts reported to the states and the states reported to the federal government had resulted in outdated and generally unreliable and noncomparable data. The NIE survey was based on a nationally representative sample of 100 school districts. Since unreliability was a feature of past studies, interviews were conducted not only with the district directors of the programs, but also with principals, teachers, and parents within each district. Interviews ranged from 30 minutes to two and a half hours. Previous studies had relied on self-report, mailed questionnaires, usually to district administrators.

The advisability of such expensive, laborious, face-to-face interviewing may be illustrated by the following anecdote. Between 1967 and 1971, I directed a large-scale evaluation of the Illinois Gifted Program, a categorical program that was funded similarly to Title I. In the initial data collection, questionnaires were mailed to all 1,100 Illinois school districts; a respectable return was received. The questionnaire asked for detailed descriptions of programs for the gifted supported by state funds. We analyzed and sent the results to the State Department of Education and the State Legisla-

ture -- we were obliged to provide a report the first year.

In the third year of the evaluation, we examined the financial claims of the school districts and found that 340 districts were receiving state funds. From these we selected a 10 percent random sample stratified by type and size of school district, and sent teams of people trained with special interview schedules, observation protocols, and psychometric instruments into each district. Generally, there was little resemblance between the original questionnaire description of the local program and what our data-collection teams found.

A few administrators admitted forthrightly that their gifted program was a fabrication which existed only on paper. In other cases, the respondents had answered sincerely, but in fact had little idea of what their programs consisted. Only when we interviewed the teachers did we discover what was happening. In even more cases, people were misled by our questionnaire. For example, when they marked "individualized instruction" as a type of program, they often meant something quite different from what we had supposed by the label. We fooled ourselves by our terminology. All of this misinformation was discovered only when interviewing several people and observing the classes within the districts.

Reliable information is often bought only at high cost. In most cases mailed questionnaires are inadequate as information sources. The chances for intentional and unintentional misinforming are too great. Fortunately, in-depth data collection was what the NIE survey undertook, which makes this part of the study seem accurate and credible. They also conducted case studies in 18 of the districts (as did we) to determine how district auxiliary services were provided. The only fault I can find with this part of the study is that they apparently did not visit classes.

Although federal money was awarded to school districts on the basis of income level of the children's parents, children became eligible for Title I services at the school level by being identified as having an educational need. They were chosen by low test scores and teacher judgment. Of those eligible, 66 percent received special educational services.

The provision of special services was directly proportional to the amount of special funding available to the district. The more money available, the more students served. There was no significant variation in student participation rates in districts of different sizes, different regions, or different socio-economic characteristics. Children in non-public schools received fewer services. Of the students

served, 59 percent were white (compared to 75 percent total enrollment), 34 percent were black (compared to 19 percent of total enrollment), and 10 percent were Spanish-surnamed (compared to 5 percent of total enrollment). Contrary to popular conception, Title I is not solely, exclusively, or even primarily for ethnic minorities. It is for the poor.

Services Rendered[6]

What kinds of services were being offered to children identified as needy? Describing educational programs is difficult because there are so many factors that may be significant and contribute to or detract from program quality. An experienced educator ordinarily has little difficulty recognizing a good or a poor class, but measuring characteristics that are indicative of high quality by standard measurement procedures is a problem.

The difficulty is that one might expect a small-size class to be a better one, all other things being equal. But other things are not usually equal, so a small class with a bad teacher will result in poor quality education. There are so many relevant factors, and they interact so strongly with one another that some researchers have despaired of being able to predict precisely the pattern of factors that will

result in quality education (Cronbach, 1975). One strategy is to forego description of the program and simply measure the outcomes. But the outcome measures themselves, such as standardized achievement tests, are often oblique or even invalid measures of educational quality. This has been the past path to failure.

What the NIE evaluators did was to identify program characteristics that educational research has suggested are associated with high quality. This is not what I would have done, but it is not a bad strategy and is certainly preferable to attempting to rely exclusively on outcome measures. Alternatives might have been to conduct in-depth case studies or to assemble a larger number of descriptive and evaluative observations, both of which would provide a denser, thicker description of the programs. (See House, Steele, and Kerins, 1971). The NIE strategy was a good one, however.

The factors selected for the study were time spent on instruction, class size, teacher qualifications, and the extent of individualization. The NIE survey revealed that three-quarters of the Title I funding went directly to instructional programs. Ninety-five percent of the school districts offered remedial reading or language arts programs. Half had only remedial reading or language arts, while 44 percent had a math program. Almost all were elementary school

programs.

The average Title I student received 3 hours and 47 minutes of special instruction a week if the program was reading, 4 hours a week if language arts, and 3 hours a week if math. From 1/4 to 1/5 of the students' time was spent in compensatory instruction. Title I programs were deliberately supplemental and in addition to regular instruction. The average class size was 9 in remedial reading and 12 in language arts and math, compared to 27 in regular classes.

As to individualization of instruction, as defined by the NIE survey, 38 percent of the teachers reported using instructional objectives, 77 percent said they used tests, and 70 percent reported teaching to small groups or individuals. Individualization, one might add, could have been defined in several different ways.

Teachers were selected primarily because of academic training or experience in compensatory education. The average total teaching experience of compensatory teachers was similar to the national average. The most highly valued training, as reported by district personnel, was in measurement and evaluation, new techniques, new materials, and characteristics of the disadvantaged. Two-thirds of the teachers had received

special training, the average being 25 hours. Most districts provided some in-service training for teachers, usually consisting of consultants, staff meetings, and workshops. Only a small portion of the Title I budget (0.5 percent) was spent on in-service training, the average district expenditure being $600.

Most compensatory services were offered as "pull-out" programs, that is, students were removed from the regular classes and put into special ones for their compensatory instruction. About 25 percent of the students were in special classes for the entire day, a feature more prevalent in districts with many eligible students. The school districts' Title I applications for funds had low utility for predicting district practices. Most of the instructional expenses consisted of teacher salaries, and the number of teachers seemed to determine the number of students served and the total time spent in reading.

In reading programs relatively more teacher aides were employed; in language arts programs, there were more teachers with bachelors degrees, and more with masters in reading and math. Pull-out programs used more aides and teachers with masters degrees. In programs with more instructional time, there were more aides and less well-educated and less well-paid personnel.

In 12 school districts, case studies were conducted of the programs, but this information was confounded by the fact that these were the "demonstration" districts which were trying out different funds allocation procedures. They were not typical districts. A closer examination of their programs revealed that the Title I students spent 136 minutes a week in language arts compared to 122 minutes for regular students. There was a great deal of variation from district to district, however. In these districts Title I students spent the same percent of time receiving instruction as individuals (15 vs. 13 percent for regular students), more time in small groups (24 vs. 9 percent), the same time in medium groups (32 vs. 32 percent), and less time in large groups (29 vs. 46 percent). Small groups were defined as 2-5 students, medium as 6-20 students, and large groups as more than 20 students. Clearly, it was in the small group instruction that the Title I students had the advantage.

Only 4.6 percent of the total budget was spent on auxiliary services in 1975-76. The federal emphasis on basic skills had resulted in yearly declines. About 1/3 of this amount was spent on parent involvement, another 1/3 on guidance and social work, and the other 1/3 on health, food, and transportation. Health services consisted of nurses, screening, immunization, referrals, and physical exams. Health and

food costs together accounted for only one percent of the budgets. The Office of Education had emphasized that Title I be an educational rather than a welfare program.

The nature of auxiliary services was determined by OE emphasis away from health and welfare, by federal emphasis on parent-advisory committees, and by the type of evaluations that were conducted. The OE guidelines demanded quantitative measures of educational outcomes, and school districts did not know how to assess health and other auxiliary services using these measures. Opinions of parents and others were considered too subjective. Consequently, districts dropped these services in favor of services they could measure with achievement tests, according to the NIE study. This is a case of programs being changed to serve the evaluation rather than vice versa, and I would condemn this effect of evaluation more strongly than do the NIE researchers.

How good were these compensatory programs? The NIE survey provides some criteria by which the programs can be judged -- if one is aware of other educational research on class size and so on. First, most money goes into salaries for instructional personnel. More money provides more instruction for more students. In this sense, the benefits are directly proportional to the money spent.

The classes themselves are small, averaging 9 students in reading and 12 in language arts and math. From Glass and Smith's (1979) meta-analysis of the effects of class size on achievement, one could expect a standardized differential achievement of .256, equivalent to about 10 percentile points, quite a significant gain. That assumes, however, that all the students' time was spent in classes of such size. Actually, only one-quarter was so spent. For classes of 12 students, one would expect a gain of about 5 percentile points. Also the expected gain would be slightly lower than this for elementary students, which most Title I programs service, and slightly higher for secondary students. Other things being equal, tutorials of one person compared to classes of 40 students could make a difference of 30 percentile points, a difference substantially greater than that demonstrated for any other instructional technique.

The fact that students in compensatory programs are spending more time in reading, language arts, and math than are students in regular instruction is also indicative of a high quality program since more time on task is associated with increased achievement (Rosenshine, 1978). Unfortunately, the NIE data are not sufficiently detailed to permit more specific estimates of gain.

Two other factors -- individualization and pull-out procedures -- are more problematic. Individualization is variously defined, and the research on its effects is mixed. The next section of the review will demonstrate this weakness. Pull-out procedures in which the children are physically removed from the classroom for special instruction probably have some harmful effects, to be discussed in detail later.

The information on teachers provided by the NIE survey I find useless as far as judging the quality of the programs. What teachers are called makes little difference in quality. If an evaluation does not arrive at a defensible judgment of worth, it should at least strive to provide information that the reader can use to make such judgments. That teachers are selected on the basis of training and previous experience is good, but the training appears to be slight on average. One wonders whether the teachers are really prepared. Also very little was spent on in-service training. One would think that there would be far more attention to improving teacher skills in such an important and huge program. An average annual district expenditure of $600 leaves much to be desired.

Finally, the federal de-emphasis on auxiliary services is acceptable <u>if</u> children are provided medical services, referrals, etc., from other sources. Correction for near-sightedness

is likely to result in far mor educational achievement than most things one can think of. Especially since medical services are so expensive, one wonders whether the children are receiving basic physical care. The Title I purposes include concern about comprehensive development of the child. The evaluation should have investigated why the Office of Education has de-emphasized auxiliary services and whether the children were being taken care of elsewhere. This is one of the serious deficiencies of the study.

To be fully assured that the compensatory programs are of high quality, one would have to investigate the classrooms in more depth. This was only one aspect of the study, and one cannot do everything. Yet, one longs for a deeper understanding of the compensatory classrooms, the "complete understanding" aspired to by the evaluators themselves. Nonetheless, I believe the NIE evaluators had their priorities essentially correct as to where to put their resources.

Student Development[7]

The third fundamental purpose of compensatory education, according to NIE, was to contribute to the cognitive, emotional, social, or physical development of the participating students. The NIE substudies which addressed this

purpose were more typical of the way such evaluations are usually done, and less satisfactory than the rest of the NIE study. The evaluators rejected the notion of a grand national summary evaluation of Title I because the past evaluations had been failures. These failures had occurred, they said, because of the diverse nature of the programs, the masking effects of a grand summary measure, and the limitations of measurement instruments. Instructional programs are not unique, nonoverlapping treatments. This was good thinking on their part.

Instead, they chose a sample of districts to investigate. The sample was selected to provide a broad array of instructional variables that could be related to achievement gains. Generally, studies that have attempted to relate quantified instructional variables to achievement gains have not fared well. The main variables here were instructional time and content, amount of individualized instruction, and the difference between instruction in regular classes and pull-out classes -- classes in which Title I students are withdrawn from the regular classroom for instruction.

The Comprehensive Test of Basic Skills was administered to students in 400 classrooms in 14 different districts in the fall, spring, and following fall. The sample was not chosen to be representative of Title I classes as a whole, so

the results cannot be generalized to all Title I classes. The investigators did claim considerable diversity of programs in the sample.

From fall to spring, a period of seven months, students in the first grade gained 12 grade equivalent months in reading and 11 months in math. Third graders gained 8 months in reading and 12 months in math. The average pupil in the norm group gained 7 months in the same time. These gains were quite high compared to other evaluation studies of compensatory education. A follow-up study in the fall found the students holding these gains on average, a finding again contrary to expectations.

One must wonder about the sample of students and the test. On the pretest the first graders averaged at the 22nd percentile in both reading and math, and the third graders averaged at the 13th percentile in reading and the 5th percentile in math -- two quite different groups of students. Besides having no experimental controls, the same tests were administered to the students on three occasions, providing opportunity to learn the test items. The results are strikingly positive but not convincing as to the effects of Title I programs.

Glass and Smith (1977) questioned these results. The pretest raw score mean for first-grade reading was 2.38, close to the chance raw

score of 24.8. The first-grade math raw score mean was 15.6, about three points below the expected mean of 18.7 if students guessed randomly. This raises the specter that Level B of the Comprehensive Test of Basic Skills was too difficult to use as a pretest for the first graders, and that the achievement gains were spuriously high. This could account for the unusually positive results. This suspicion is strengthened by the low pretest and post-test correlations on the same test for first grade ($r = .50$ in reading and .39 in math compared to .86 and .78 in third grade).

As in most such studies, the focus of measurement was narrow -- only reading and math at a few grade levels. The evaluators felt that there was inadequate instrumentation to assess the other areas of student development. In this they were probably correct, but nonetheless judgments about the program were made on this narrow basis. Interviews with students, parents, and teachers would have been valuable in assessing development. The researchers report in a footnote that student attitudes were high in the fall and the spring but give no further information about this aspect of the study.

How these instructional variables related to gains in test scores was also ambiguous. The first graders gained more in the regular classrooms, in both reading and math, but the third

graders in the pull-out classes did better in math and the same in reading. Was this because the students were different or because of spurious measurement? One thing is clear: student gains were large when the instruction emphasized the skills measured by the test. It was possible to measure the overlap with the curricula by analyzing the test items and the video tapes of classes, instructional materials, and interviews with teachers. This was an important precaution that is seldom taken, and one of the best features of the study.

Classroom processes were measured by indicators of opportunity to learn, individualization, "instructional events," motivators, and teacher background. Taken as a whole, these classroom processes accounted for a significant amount (.14 to .27) of the variance in achievement gains in first and third grades. In this analysis the evaluators quite properly used the classroom rather than the individual student as the unit of analysis. These figures and others are taken from a later and more detailed commentary than that provided in the original report (Cooley and Leinhardt, 1980).

When the five classroom process indices were taken separately and in combination, only a few accounted for modest proportions of the achievement gain. "Opportunity" accounted for .22 of the variance in first grade reading, .09

in math, .09 in third grade reading, and .13 in third grade math. Most contributions were in the range of .01 to .04 of the variance in test gain.

When one looks more closely at the "opportunity to learn" index, it is dominated by the overlap measure between the test and curriculum. This is a combination of two scores, one an estimate by teachers of whether they had taught the test items and the other based on an analysis of the curriculum by experts. Zero-order correlations of this variable with the test gain were .40, .25, .29, and .38 (in reading and math at first and third grade). Clearly, test overlap accounted for a substantial portion of the "opportunity" index. Furthermore, when one examines all the other single variables, the highest correlations were with "time opportunity," "student mastery," and "matching" in variables. The latter two measure test-taking practice in the curriculum.

The investigators concluded that what is taught is more important (in affecting test scores) than how it is taught. Children will do much better on tests if they are exposed to both the form and content of the test itself. The researchers recommended an intensive examination of the match between the content of curricula and tests. Future evaluation, they suggested, should contain a measure of curriculum fit to

the test.

I agree with this recommendation but would turn their suggestion around. The most parsimonious explanation of these results is not that the curricula are in error somehow, but that the tests are. They are far too narrow in scope. If one insists on evaluating curricula with narrowly focused tests, the results will largely reflect the degree to which the curriculum teaches these particular test items. One can indeed develop a curriculum to teach those test items and do better on the tests, but such education would be strange indeed, and perhaps worthless. It is an inverted logic which insists that schools teach what standardized tests measure. One can envision an entire society giving itself tests and succeeding on them as it becomes progressively removed from reality and experience.

The main finding of the study, as seen by the investigators, was that classes with individualized instruction (as defined in the study) were no better than those without. The investigators expressed the opinion that there was a low degree of individualization in most classes, in their opinion. Apparently, the Congress was considering mandating such techniques in 1974 and was deterred by the lack of positive evidence about individualization here and elsewhere. This is another mandate that we are far

better off without, and the study was certainly valuable if it helped prevent such an action.

Overall, in spite of some interesting ideas about how to measure and analyze instructional variables in a way closer to classroom experience, the study did not provide a very full or convincing picture of student development in compensatory classes. Gains in the classroom seemed to be related to certain activities, but one can't say much more. The study becomes one more equivocal essay of its type. It is more suggestive than persuasive. The results are not satisfying.

Because of the flaws, limitations, and mixed results, the NIE study is also inconclusive about pull-out programs. This is an important issue for compensatory programs because the majority of Title I students received instruction in pull-out classes. Fortunately, the Office of Education commissioned a policy study by Glass and Smith (1977) on the issue, which proved to be more persuasive. Although this study is not part of the NIE study, the issue is so important for Title I that I mention it here.

To address the issue, Glass and Smith conducted an analysis and meta-analysis of research studies from related areas such as ability grouping, mainstreaming, and desegregation. Combining these studies, they found that labeling

students as "slow" or "weak" reduced their academic performance by one-quarter standard deviation below that for comparable pupils not so laeled. Teacher attention and support for the pupils was reduced by one-third standard deviation and teacher judgment of student success by a half standard deviation. From their analysis "the pull-out procedure per se has no clear academic or social benefits and may, in fact, be detrimental to pupils' progress and adjustment to school." (Glass and Smith, 1977, p. 7). This study is more persuasive because it relies on the combination of many strands of research rather than attempting to base its conclusions on one study with all its idiosyncracies and weaknesses. Also, in this case the findings from various areas are consistent in indicating that the pull-out procedure might be detrimental. The dependent measures are all cognitive, however.

Glass and Smith took the position that I advocated earlier: all pupils in poor schools deserve compensatory services. Any child in a school with a high concentration of poor pupils is put at an educational disadvantage. By aiming funds at schools with large concentrations of poor children, the beneficial effects of categorical aid could be retained and the invidious effects of pull-out procedures avoided. Only a matter of administrative convenience in monitoring has led to the pull-out practice. In

fact, the management of the program would be better if the unit were the school.

This is an acute case of officials being so concerned that someone ineligible might benefit that those who rightfully should benefit are deprived, a case of monitoring working against the intent of the program. While the zealousness of concern for the poor is laudatory, one must be careful when pursuing such policies that they are not counter-productive. Professional judgment, research, and common sense argue against the pull-out procedure.

In this particular case, the OE study succeeded where the NIE study did not. However, the overall effects of these compensatory programs on student development remain unassessed. Educational measurement continues to be inadequate to the task of assessing braod student development, there are no magic solutions; evaluators should be more honest about what they cannot do.

Summary

From reviewing the NIE compensatory education study, I have two strong impressions. One is that the compensatory programs, especially Title I, are very good programs.[8] The second impression is that it is possible to evaluate

such programs intelligently, usefully, and validly. That these enterprises were the responsibility of the federal government is striking, for the federal performance on these matters has not been good over the past several years.

My impression of Title I is that it is a categorical program that has worked. The purpose was to help the children in low income districts, and they are being helped. Funds go to the regions, districts, cities, and schools with lowest incomes. Compared to other federal and state programs, it is perhaps the most redistributive of all, an impressive accomplishment.

Of course, it will not eradicate poverty. The resources are too meager for that, and probably poverty stems from quite different sources than the schools. Eradicating poverty is too much to expect. The funds are enough to make a discernible difference in the education of the children in poor areas, however. Other methods of allocation would be less equitable and further removed from the purposes of the program, as well as less accurate and less efficient.

I am persuaded by the NIE study that most of the funds allocation requirements are necessary and not overly restrictive. The program development requirements, on the other hand, edge toward over-regulation and the mandating of

practices for which there is little supporting evidence, scientific or professional. This seems to me to be an all too typical misguided foray into prescribing practices that experienced people on the scene would not endorse.

Unfortunately, many government planners cannot seem to tell the difference between requirements that are necessary because local incentives work against them, as with funds allocation, and requirements that are detrimental to sound educational practice, which is determined by experience as in any craft. The worst example is the mandating of particular types of evaluations that are contrary to the expertise of the evaluation community.

I am persuaded that the management of the compensatory program at all levels is reasonably efficient and effective, although I suspect that much of the time of state and local administrators is spent meeting requirements that have little to do with classes for children. I suspect that the real job is done by extra personnel at the school and classroom level who do what they can. Most of the funding supports these extra personnel.

The monitoring of these programs appears confused and conflicted because of the impasse on criteria for monitoring inside the Office of Education. This situation might have been par-

tially resolved, and the monitoring made easier, by making entire schools eligible in low income areas. Such a policy might have countered the apparently deleterious effects of the pull-out procedures, which were instituted for fear of violating unclear and ambiguous federal guidelines. Overall, the programs appear to be of high quality, primarily because of the employment of more personnel. They focus on the basics in reading and math. Exactly what they consist of is not clear, though one might question the effectiveness of pull-out procedures and individualized instruction.

I am not persuaded as to how students are developing in and as a result of these programs. Although the test score gains were very impressive over a one-year period in the sample classrooms, there are a number of problems with the study of student development. In addition, even if the gains were real and indicative of overall quality, the aspects of development investigated were narrowly construed. Both the exact nature of the compensatory classrooms and the specific development of children in these classes remain hazy.

For better or for worse, I have gathered these impressions primarily from the NIE study. I have had little first-had experience with these programs or classes. The persuasiveness of the NIE study is partially dependent on its

validity. It is well-conceived as a series of substudies addressing a large number of issues.

In its intensive investigation by interview, observation, and inspection, it seems to be accurate. Loose ends are pursued, alternative hypotheses examined. In other words, the study is generally conceptually well-reasoned. The possible exception here is the work on student development, which leaves a number of questions unanswered. There are too many flaws for these conclusions to be persuasive.

The NIE study as a whole is also credible. Its enthusiastic reception by Congress is one indicator of its relevance and credibility to its major audience. It was done by an agency external to the program and by different researchers and groups, many of them well known. The results are also internally consistent. The final report could have been more coherent by drawing the many disparate strands together.

Finally, the evaluation was just and fair and democratic. A just evaluation should attend to the interests of all groups, particularly to those who are most affected by the programs being examined. In this case in particular the study should attend to the interests of the poor and disadvantaged. The NIE study accomplished this by examining in great detail how the resources and services were distributed. Justice

is concerned with distribution, with who gets what. Through examination of the program administration, the investigators also determined how recipients and participants were treated. Consistency and the quality of these relationships were important concerns. Of course, in hindsight, all of these things could have been improved, but in my opinion they were remarkably well handled in this study.

I believe the foregoing fundamental considerations -- the truth, credibility, and correctness of the study -- account for its validity. Validity in turn means that the evaluation is acceptable and that the audience, Congress in this case, can act upon it. Of course, Congress may not act upon it. There may be reasons, legitimate or otherwise, that prevent Congress from acting.

Because an evaluation is valid and acceptable does not mean that it is accepted. Perhaps key Congressional leaders know that another fund allocation scheme would benefit their own districts more than the scheme favored in the study. The self-interest of their district might override their concern for the poor. Or perhaps some important factor changes while the study is in progress -- the funds disappear. Or perhaps the study is lost someplace in the vagaries of the legislative process. For various reasons the study may be valid and acceptable

but not used.

If the study is used, however, there is normally a presumption of its validity. Certainly the Congress does not examine the implicit validity claims as I have done. Rather, as expressed earlier, its concern is not with technicalities but with whether the information can withstand a professional critique and give it some guidance as to what to do. The Congress presumes validity.

It might also be in highly politicized situations that people cynically use only that information which will benefit their cause. Even in this circumstance, however, the information is only effective if the people on whom it is used believe that it is valid. It is not enough that an evaluation be used; it must be valid. And ordinarily it will not be used unless it is at least presumed, rightly or wrongly, to be valid.

In this case there was no problem. The NIE study was valid, according to standards I have enunciated, and it was used by the Congress to very good effect. In this case it served the Congress well and was a good evaluation from both a Congressional and a professional evaluator's point of view.

FOOTNOTES

1. Hill pays special tribute to the help of David Wiley and Jim Vanecko in the conduct of the study.

2. Unless otherwise noted, all facts and figures about Title I are taken from the NIE study, which was published in seven separate volumes. These figures are from Volume 2.

3. The information in this section is primarily from Volumes 3 and 4.

4. Congress did eventually reject the notion of allocation by test score. One factor, according to Hill, was that the study did not permit the estimation of allocations for individual Congressional districts. Not knowing what their districts would receive under a new allocation scheme, Congressmen retained the old allocation procedure.

5. The information in this section is from Volume 5.

6. The information in this section is primarily from Volume I.

7. The information in this section is primarily from Volumes 6 and 7.

8. The Title I programs have undergone substantial changes since these studies were begun in 1974, partly as a result of these studies. I am not well informed enough about the current program to suggest exactly what these changes have been.

REFERENCES

Cooley, W. W., & Leinhardt, G. The instructional dimensions study. Educational Evaluation and Policy Analysis, 1980, 2(1), 7-25.

Cronbach, L. J. Beyond the two disciplines of scientific psychology. American Psychologist, 1975, 30, 116-127.

Cross, C. T. Title I evaluation -- a case study in congressional frustration. Educational Evaluation and Policy Analysis, 1979, 1(2), 15-21.

Glass, G. V., & Smith, M. L. Meta-analysis of research on class size and achievement. Educational Evaluation and Policy Analysis, 1979, 1(1), 2-15.

Glass, G. V., & Smith, M. L. "Pull Out" in compensatory education. Prepared for the Office of the Commissioner, U. S. Office of Education, November 2, 1977.

House, E. R. Evaluating with validity. Beverly Hills, CA: SAGE Publications, 1980.

House, E. R., Steele, J. M., & Kerins, C. T. The gifted classroom. Urbana: Center for Instructional Research and Curriculum Evaluation, University of Illinois, 1971.

Jaeger, R. M. The effect of test selection on Title I project impact. Educational Evaluation and Policy Analysis, 1979, 1(2), 33-40.

Lindblom, C. E., & Cohen, D. K. Usable knowledge. New Haven, CT: Yale University Press, 1979.

Linn, R. L. Validity of inferences based on the proposed Title I evaluation models. Educational Evaluation and Policy Analysis, 1979, 1(2), 23-32.

McLaughlin, M. W. Evaluation and reform. Cambridge, MA: Ballinger Publishing, 1975.

Rosenshine, B. V. Academic engaged time, content covered, and direct instruction. Journal of Education, 1978, 160(3), 38-66.

Roth, J. Needs and the needs assessment process. Evaluation News, 1977, No. 3, pp. 15-17.

Scriven, M., & Roth, J. Needs assessment. Evaluation News, 1977, No. 2, pp. 25-28.

Wiley, D. E. Evaluation by aggregation: social and methodological biases. Educational Evaluation and Policy Analysis, 1979, 1(2), 41-45.

Evaluating Compensatory Education

Torsten Husén, University of Stockholm

Particulars About the Point of Departure

I have in my office four big files of newspaper clippings from the New York Times, the San Francisco Chronicle, and the Palo Alto Times on educational issues at the national, state, and local level collected during the academic year 1965-66 which I spent at the Stanford Center for Advanced Study in the Behavioral Sciences. I tried to take advantage of my year at this "think-tank" to deepen my knowledge of American education, a process that had started more than ten years earlier when I visited half a dozen universities and later spent some time as a visiting professor at the Graduate School of Education at the University of Chicago. One of my thick files contains clippings from 1965-66 on Head Start, urban education, segregation, vandalism, education of the minorities, etc., in short, social-pedagogical problems. The articles emanate from the heydays of the new legislation of the Johnson administration on compensatory education passed under the auspices of the Great Society.

I did not envisage, when this material was collected, that it would prove useful in provid-

ing a time perspective for the review of U.S. federal policies for the education of the disadvantaged launched by the Organization for Economic Cooperation and Development (OECD) in Paris as one of the so-called country reviews on educational policies that OECD sponsors in cooperation with the government whose policies are under scrutiny. Nor did I envisage that I should be serving as the rapporteur of the review team which implied that several additional months had to be spent both before and after the site visits on covering, hopefully, relevant documents and research literature and, of course, writing the report. The latter is now available in print (OECD, 1980).

The present review has been commissioned to focus on the NIE reports requested by Congress in 1974. I shall, however, view the reports in the wider perspective provided by, among other things, the experiences gained during the OECD policy review. Reference will also be made to some of the 35 projects which form the basis for the six NIE reports listed above and which in addition to the interim reports form the basis for the final report on compensatory education that NIE submitted in September 1978 shortly before Congress was about to take final decisions on a 5-year re-authorization of the various programs, and in particular about those under ESEA, Title I.

Since the OECD policy review provides the framework for my attempts to review the NIE studies of compensatory education programs, it would be in order to give a brief account of how this exercise was carried out. In the case of the United States it was on the whole conducted according to the same procedure as other OECD country reviews with the important difference that it was limited to the education of the disadvantaged. Two previous country reviews, France in 1969-70 and the Federal Republic of Germany in 1971-72 (OECD, 1971 and 1972), in which I also participated, covered the entire field of national educational policies. The following steps can be distinguished in a country review procedure:

1. The initiative to conduct a policy review can be taken by the government of a particular country which wants its policies to be critically scrutinized by outside people not entangled in its own internal political controversies and who are free to raise issues which can not easily be raised within the country. The OECD secretariat can also initiate a review of policies in a particular country, which is considered to be of common interest and benefit to the member countries. Once the decision has been taken to conduct a review, the OECD secretariat in contact with the Government concerned appoints a team of reviewers, as a rule persons who are not only knowledgeable about the

educational systems of the country but also carry a reputation in the international community of scholars that lends sufficient weight to their work.

2. The government under review prepares a National Report which essentially is an attempt to collate relevant legislative, statistical, and research information pertaining to its educational system. In the case of the U.S. review of compensatory education the U.S. Office of Education (OE, 1979) prepared six extensive case studies of compensatory education programs, such as Title I, the bi-lingual programs, the basic Educational Opportunity Grant program, etc. The reviewers received in advance piles of legislative, executive, and research documents. The task even to cover the most relevant ones was enormous, particularly for the rapporteur who in the U.S. case, assisted by Dr. Lawrence Saha of the Australian National University, had to prepare an issues paper. A few examples will illustrate the gigantic documentation problem. The review of Head Start research since 1969 and the bibliography that goes with it commissioned by HEW and published in 1978 is a 158-page document (Mann et al., 1977). The Abt Associates, Inc. evaluation of the Follow Through Planned Experiments is reported in seven volumes, etc. In addition, HEW through the Office of Education submits a voluminous annual evaluation report to Congress on its programs.

3. Site visits and interviews with key people in the country under review are conducted by the entire review team. In the case of the U.S. the team spent about one week interviewing people in Washington, D.C., beginning in the White House, proceeding with the Office of Education and other Federal agencies and with the relevant Congressional committees. It also tried to meet representatives, as many as possible, of the countless "advocacy groups" that are based in Washington, D.C. We then traveled to 15 different places in nine states and had a wrap-up session in Washington, D.C. This was what we were able to achieve in a 3-4-week period. The rapporteur spent another four months working on the report, being during that period based at the National Humanities Center in Research Triangle Park, North Carolina.

4. The reviewers at the wrap-up session decided on a series of issues that they wanted to raise with the representatives of the Federal government. After a few months a preliminary report was made available to the partners involved along with a list of major issues broken down in specific questions. The report and the section on issues formed the background document for a "confrontation meeting" at the OECD headquarters in Paris under the auspices and presence of the OECD Education Committee consisting of representatives of the 24 member countries. The discussion as usual took place in a kind of

seminar setting with one reviewer introducing one major issue or question at a time, whereupon in this case the U.S. team representing the Federal government responded. Members of the Education Committee can and actually did also take the floor and posed questions to the responding team. The report, together with an account of the proceedings at the confrontation meeting, is then printed in English and French.

Again, in preparing the present review of the NIE reports on compensatory education I have quite naturally drawn heavily on the parts of the OECD report which I wrote up. The more technical aspects of the evaluation studies were for the most part not considered in the OECD report but have of course been included here.

The NIE Evaluation Exercise

In connection with the passing of the 1974 ESEA Amendments, the 93rd Congress requested NIE to conduct a comprehensive study of the compensatory education programs. The Institute was instructed by Congress to "undertake a thorough evaluation and study of compensatory education programs, including such programs conducted by States and such programs conducted under Title I of the Elementary and Secondary Education Act of 1965." The Institute was specifically instructed to conduct

"(1) an examination of the fundamental purposes of such programs, and the effectiveness of such programs in attaining such purposes;

(2) an analysis of means to identify accurately the children who have the greatest needs for such programs, in keeping with the fundamental purposes thereof;

(3) an analysis of the effectiveness of the methods and procedures for meeting the educational needs of children, including the use of individualized written educational plans for children, and for programs for training the teachers of children;

(4) an exploration of alternative methods, including the use of procedures to assess educational disadvantage, for distributing funds under such programs to States, to State educational agencies, and to local educational agencies in an equitable and efficient manner, which will accurately reflect current conditions and insure that such funds reach the areas of greatest need and are effectively used for such areas..." (NIE, 1977f, p. 42).

In order to achieve the purposes under the four clauses cited the Institute was instructed to carry out "not more than 20 experimental programs" which could be commissioned to agencies eligible for grants.

NIE in accordance with the legislative provision in 1977 submitted six interim reports (NIE, 1977 a-f) and a Final Report (NIE, 1978) to Congress in September 1978 shortly before the Education Admendments were on the floor.

This was, indeed, an overwhelming charge typical of the aspirations held by policy makers when turning to researchers. As I have spelled out in another connection (Husén, 1978), the problems as formulated by policy makers must be broken down in manageable units in order to become researchable. Therefore, NIE apparently has tried to narrow down the package of tasks without deviating too much from the letter of the law. It has attempted to restrict its charge in two respects. "In response to this request, NIE implemented a study of compensatory education programs, focusing on Title I of the Elementary and Secondary Act of 1965", is the laconic interpretation that NIE makes in the introductory section of its final evaluation report (NIE, 1978). But within Title I there was a limitation to "four major study areas": funds allocation, service delivery, student development, and program administration, all technical in nature and amenable to down-to-earth empirical studies. One should, for reasons just given, not criticize NIE for opting for such limitations. Already the task it decided to undertake was enormous. Notwithstanding this, one nevertheless lacks the wider perspective on the

compensatory education programs that the OECD review team, given its wider terms of reference, was free to adopt. Thus, in the reports one does not find any considerable concern with the "fundamental purposes" envisaged by the Law. Apart from the immanent difficulties stemming from the breadth of such a task, one should consider its political sensitivity. To question basic purposes and strategies is to place oneself in the middle of political controversy. The limitation to Title I can, of course, be defended on the ground that the overwhelming portion of the compensatory education funds goes into Title I programs.

The six interim reports, as well as the major portion of the Final Report, are based on 35 research projects conducted for NIE by various research organizations, such as Abt Associates, Inc., National Opinion Research Center, and the Syracuse Research Corporation.

Even given the considerable restrictions the NIE has imposed on itself in conducting the Compensatory Education Study under congressional instruction, its reports cover a wide range of problems pertaining to program administration, service delivery, fund allocation, and student development. In reviewing the study I shall not focus on technicalities but take the liberty of applying the same broad, policy-oriented perspective as did the OECD team and thereby draw

heavily on my experiences from this review as well as on the material I prepared as rapporteur. Considering the audience for which the OECD report was prepared, the technical and scholarly aspects of the execution of the Compensatory Education Study were dealt with briefly and cursorily only. In the present review they take a more prominent place. However, it would be highly presumptuous even within the space allotted to the present review to claim a thorough critique focusing on the massive efforts made in the various commissioned projects launched to evaluate specific aspects of the Title I programs. In addition to the 35 projects, on which the NIE Compensatory Education Study in the first place has been based, there is an enormous scholarly literature pertaining to compensatory education programs of various kinds.

I shall not review the NIE reports one by one but make casual references as I move along. After having "established my credentials", by indicating what kind of background I bring to the review task, I shall as a background present the main features of the compensatory education legislation. A large portion of my review will focus on what the legislators referred to as the "fundamental purposes", that is to say, the philosophy and strategies behind compensatory education. On the basis of such an analysis I shall discuss the targeting of Title I funds,

both its horizontal and vertical coverage. Finally, I shall deal with the massive evaluation efforts that have been conducted particularly of the Follow-Through Planned Variation Experiments.

I have already mentioned the conspicuous lack of studies of the "fundamental purposes" and hastened to indicate that this is something that a government agency, such as the NIE, could not easily deal with. It would rather have been the task for an institution, such as the National Academy of Education. Within the restrictions that NIE has set for its research efforts and -- not least -- within the limits of what it could reasonably be expected to achieve within the given time-frame and resources, its Compensatory Education Study has considerable merits. One has succeeded in collecting relevant research either already conducted or commissioned to outside agencies and in conducting in-house studies. All these together present what is essentially a good descriptive picture of how Title I compensatory education is operating and what its main technical and administrative problems are.

The report (NIE, 1977 f) on the use of achievement test scores to allocate Title I funds presents a sound empirical basis as well as a good analysis of the technical problems involved in switching from census-based poverty

criteria to achievement criteria. Not least the financial implications, particularly important to legislators, have been well spelled out. It is, however, difficult to avoid the impression that the difficulties that have to be overcome in a switch to achievement criteria have been given stronger emphasis than the disadvantages of the present system. The mistargeting and its conceptual determinants had deserved more thorough studies. Since congressmen eagerly watch how changes in allocation criteria affect their constituencies, findings that too conspicuously reveal inadequacies of present practices are easily felt as stepping on their toes. The vested interests in status quo in a multibillion enterprise, such as compensatory education programs, are indeed strong. Congress did in its 1978 Education Amendments not make any considerable change in previous practices of allocating funds. The only major changes enacted were those dealing with the administrative procedures, which were simplified, hopefully on the basis of findings from the NIE study of administration of compensatory education (NIE, 1977 a).

Strategies for Compensatory Education

The role Americans traditionally have assigned to education is that of a vehicle for upward social mobility. Education has more

strongly than in any other country been conceived as an instrument of realizing the American Dream. Coupled with the strong belief in education as a promotor of individual careers has been the confidence in it as a means of solving social problems by making people more competent and knowledgeable. Horace Mann, one of the architects of the American public school, conceived this school as a tool of alleviating inequalities and the disadvantages of the working classes. He wrote in 1848: "Education.... beyond all other devices of human origin is the great equalizer of the conditions of men -- the balance wheel of the social machinery." (quoted after Hechinger, 1976). The confidence in education as a catalyst in changing society was behind the progressive movement in education between the two wars. Thus, at the core of George Count's reconstructivism was the concept of education as an instrument in bringing about social reform.

The belief in education as a means of reducing and, in the long run, eliminating social inequalities was an underlying force behind the Great Society legislation. The President's Task Force on Education, chaired by John W. Gardner, seems to have played an important role in preparing subsequent legislative programs for compensatory education of the disadvantaged. Not only did the Task Force confirm an adherence to the traditional tenet by saying that it would be

a major task for the Nation to "provide every child with as much education as his talent and drive warrant", but it also recommended steps to be taken to give children of disadvantaged background access to normal educational facilities. The Task Force pointed out the great inequalities between the States and localities in paying for education and even recommended general Federal aid to school education on the basis of some equalization formula, taking, however, a dim view of its "political feasibility" (Report, 1964).

The Presidential Task Force pointed out that the American school so far had done well with children who belonged to the mainstream but had neglected those at both ends of the spectrum, the exceptionally talented at one end and the poor and physically or mentally handicapped at the other end. It pointed out that the handicapped were left out of the picture with regard to Federal support and that it would be a Federal obligation to take care of the "poor children...(who)...are to be found in our rural and urban slums, and these slums breed conditions that do in fact diminish the teachability of the child" (ibid.).

The Task Force recommended two educational strategies by means of which one should try to come to grips with the poverty problem. First, since children from poverty areas already at

school entry had fallen behind their age mates, experiments with pre-school programs would provide stimulation that could make up for environmental deficiencies. Secondly, work-training programs could be launched at the high school level.

The Federal government's commitment to what was called compensatory education began with the passing of the Vocational Education Act of 1963 and the Economic Opportunity Act of 1964. The Civil Rights movement rapidly gained momentum in the early 1960's at a time when social scientists began to explore the "other America". The Head Start program focusing on pre-school children in the form of enrichment programs was part of the overall anti-poverty program.

The big breakthrough of Federal commitments to the education of the disadvantaged was the passing of the Elementary and Secondary Education Act (ESEA) of 1965. In signing it, the President commented: "No law I have signed or will ever sign means more to the future of America." High hopes were indeed held about what education could do in rectifying injustices and inequalities in American society. The legislation embodied in ESEA could also be regarded as a breakthrough for "categorical" Federal aid on a large scale to local school agencies, although the National Defense Education Act of 1958, aiming at upgrading science, mathematics and tech-

nology, presaged further major thrusts. Since the American Constitution lacks specifications of Federal responsibility for education, it has by default been concluded that support of public education is a prerogative of the States and the localities. The role of the Federal Government is viewed as that of a partner with the States and local agencies in supporting clearly defined purposes considered to be of great importance to the nation as a whole, such as education for national defense, improving vocational education, equalization of access to higher education, and the education of the disadvantaged.

The strong emphasis put on the categorical nature of the Federal aid that it should "supplement and not supplant" explains the meticulous measures taken in order to ascertain that the States and localities did not use Federal funds as a source of general support of school education. This in turn accounts for the enormous amount of bureaucratization that has gone into the system and the resistance on its part to undergo changes, for instance, switch from economic to achievement criteria in allocating funds.

Objectives of Compensatory Education.

In instructing the NIE to "undertake a thorough evaluation and study of compensatory

education programs" it also called for "an examination of the fundamental purposes of such programs" (NIE, 1977 f, p. 42). The Institute wisely ducked the responsibility to investigate the "fundamental purposes" which, to be sure, are politically highly sensitive. The central issue here is simply how tenable the rationale behind the legislation is. Both the OECD team in general, being charged to conduct a policy review, and the present author are more protected when questioning certain assumptions behind the compensatory education legislation.

In his State of the Union message to the Congress in 1965 President Johnson had a section on aid to low-income school districts which he suggested should be given high priority. He pointed out that lack of formal education means "lower wages, frequent unemployment, and a home in an urban and rural slum." He added: "Poverty has many roots, but the taproot is ignorance." Francis Keppel, at that time U.S. Commissioner of Education, shared the conviction about the role formal education could play in reducing poverty and increasing life-chances of the offspring of the poor. He turned against the "myth" that "slum children were somehow of a lower social order of capacity than children in other sectors of our economy." Labels, such as "culturally deprived" and "socially disadvantaged" easily become "alibis for failure to find effective ways to educate these children."

(Keppel, 1966, p. 39).

Speaking of labels, the term "disadvantaged" appears since the mid-60's to be the one most frequently used in referring to children served by programs providing in the words of the ESEA "financial assistance to meet the special needs of children of low-income families and the impact that concentrations of low-income families have in the ability of local educational agencies to support adequate educational programs." (H. R. Report No. 95-1753, p. 11-12). However, one finds that the terms "underprivileged" and "deprived" are often used more or less synonymously with "disadvantaged." A typology suggested by Martin Trow (1978) appears to be useful in making a proper distinction between "disadvantaged" and "deprived." Trow classifies children according to two main dimensions: (1) material and financial resources of the family needed to enable the children to develop their potential, and (2) adequacy of socialization and education effecting the children at home and in school. The two dimensions generate a typology, where one could distinguish between four groups: (1) advantaged children with both adequate material background and socialization, (2) disadvantaged children with adequate socialization but inadequate material resources, (3) alienated children with adequate material resources but inadequate socialization, and (4) deprived children with both inadequate financial re-

sources and socialization. The important distinction between the second and the fourth group has not always been clearly made, sometimes with serious consequences for programs of social intervention. Such a distinction is of particular importance in dealing with young people during the later years in high school and in transition to working life. The reason some of them cannot hold jobs after completion of schooling is not just lack of cognitive skills, which was the main belief expressed when major compensatory education programs for teenagers were drawn up. Deficient socialization bringing about lack of motivation and reliability is of equal importance.

Briefly, the rationale behind the compensatory education programs was -- and among many legislators still is -- the following. Failure and/or underachievement in school is mainly due to poverty. The poor tend to aggregate in certain areas, "slums" or "poverty pockets", particularly in urban ghettos. In order to alleviate their plight and provide them with equal opportunity to achieve on par with their materially more priviledged schoolmates they should be given extra assistance. The desired outcomes of such programs of special aid should be more equal life chances among children from all walks of life in American society.

Critique of the Rationale Behind Compensatory Education.

The logic behind compensatory education could be summarized in a simple syllogism presented in the OECD Report (OECD, 1980):

1. Poverty and school achievements are closely linked.

2. Social mobility and life chances are closely linked to educational attainments.

3. Ergo: Concentrated efforts by way of increased school resources, purchased by increased financial resources targeted on disadvantaged students, will "break the poverty cycle."

When President Johnson declared his War on Poverty he is quoted to have said: "We are going to eliminate poverty by education....people are going to learn their way out of poverty."

The two basic propositions on a close relationship between poverty and educational failure as well as between educational attainment and life chances can on the basis of empirical evidence be questioned. Doubts raised about the propositions, of course, challenge the validity of the conclusion.

Surveys, among them the Title I programs, indicate that material resources in the home, as measured by family income, are weakly correlated with educational achievement at the level of a 0.3 correlation coefficient (NIE, 1977e, cf. Husén, 1975). This means that the overwhelming portion of variations in school achievements is accounted for by non-economic factors. The considerable spread at any given income level of school achievements was one reason why Congress instructed NIE to investigate the possibilities of using achievements instead of poverty measures as the basis for allocating compensatory funds.

More important, in the light of recent studies, not least the ones conducted by Jencks and his associates (Jencks et al., 1972; Jencks et al., 1979) one could further question the capability of the school as an institution to make up for the handicaps caused by material and/or cultural poverty at home. This is a pervasive issue in public policies which aims by means of social intervention to improve the life chances of young people. At issue is: to what extent can compensatory education, consisting mainly of additional teaching, targeted on economically handicapped children, serve as a substitute for reforms that would affect the social and economic order at large? After all, one does not need to be a Marxist to realize that education does not operate in a socio-economic

vacuum!

Another overriding philosophical issue, to which the NIE Reports did not address themselves, but which is germane to the "fundamental purposes" of the programs, is equality in education and in life chances, what this means, and how it can be achieved. What kind of equality is intended: equality of opportunity or equality of results? To what extent is formal equality in educational opportunities conducive to equality in life chances? Provisions of formal equality of opportunity in education does by no means guarantee equal results. Some children are, to use an Orwellian travesty, from the outset "more equal than others," for instance, by being born by parents possessing privileged status, not to speak of possessing more favourable genes.

The traditional confidence in what education can do has led Americans to the belief that by removing educational disadvantage one would bring about more social equality. At the level of sophistication where legislation was conceived, equality was of the classical, liberal brand. Equality of opportunity is traditionally conceived in terms of conditions that prevail in the student's background: his social class, ethnic group, sex, etc. None of these should be allowed to be an advantage or disadvantage in shaping a person's educational and vocational

career. Merit, that is to say, proven ability not background, should count in getting access to education and to subsequent jobs. Thus, public policy in education should try to remove the economic and other barriers that prevent those who grow up under disadvantaged conditions to obtain in the first place the formal education that matches their innate potential. The dilemma, which one has not until recently begun to recognize more clearly, is that all criteria of merit -- test scores, achievements in examinations, and school marks -- are significantly correlated with social background. This means that whatever selection or self-selection that takes place for entry to an educational institution as well as achievements in that institution is correlated with home background.

The legislative programs on compensatory education launched in the mid-1960's aimed at bringing about more equality of results, in terms both of educational and vocational attainments. This should be achieved by providing extra funds for children in poverty areas who were lagging behind in their school results. What actually could be achieved by providing extra funds was limited by two conditions. In the first place, children meeting the poverty definitions are to a large extent found in localities where all students in public schools are disadvantaged with regard to initial financial support. Extra funds to poverty-ridden

districts and schools therefore do not make up for the initial inequality. But there is also a fundamental limiting condition. In the debate stimulated by OECD programs as well as through recent empirical studies (Jencks et al., 1972; Sewell et al., 1976; Jencks et al., 1979) it has been shown that equality of opportunity brought about by formal equality in access and removal of economic barriers is no guarantee that other forms of equality will follow. Programs designed to ameliorate, or even eliminate, obstacles against access to further education are not enough in overcoming inequalities. Again, some are from the outset, not least by family background, "more equal than others."

Compensatory Education Programs: A Brief Overview.

I shall not here try to give a detailed description of the Federal programs for the education of poor and disadvantaged students in America. The National Report (OE, 1979) prepared for the OECD review identified some 50 programs serving the educationally disadvantaged and in six excellent case studies dealt with ten of them in detail. The NIE evaluation studies are, as pointed out above, almost entirely limited to the Title I program which is absorbing some 80 percent of funds going into compensatory education and therefore deserves to be treated

as if it were the only major program. It should, however, be pointed out that the Head Start program is still operating, with funds that are, if not absolutely so, relatively heavily reduced in comparison with the resources ten years ago. Among the compensatory programs in elementary and secondary school the Title VII program, bi-lingual education, is of pivotal importance for Hispanics, who tend to become the biggest minority in America. The National Report deals with two programs in higher education, namely the Basic Educational Opportunity Grants program and the Special Services for Disadvantaged Students. Given the scope of my assignment to review the NIE reports on compensatory education, which have by and large been limited to Title I, I shall also limit myself to the Title I program, originally launched within the framework of the ESEA.

The by far largest compensatory education effort under Federal auspices has been the Title I of the ESEA passed by Congress in 1965. Title I provided funds for special educational efforts in schools located in areas with high concentration of poor students according to four criteria: (1) number of children aged 5 to 17 in a given county who are below the Orshansky poverty line, (2) two-thirds of the children in the same age bracket who receive payments under Aid to Families with Dependent Children (AFDC), (3) children in the same age bracket who live in

special institutions, and (4) the average level of school expenditure in the State.

The amount of funds going into Title I programs has been substantial, not to say staggering. During the 1968-78 period Congress allocated 18.4 billion dollars. By 1975 Title I funding comprised 84 percent of the total ESEA funding. By late 1970's the funding level had risen to 2.5 billion from 1.2 billion ten years earlier. Title I alone accounts for about 3-4 percent of the total national expenditures for public elementary and secondary education.

The NIE study on the efficiency of Title I found that by 1976 90 percent of the school districts received Title I funds (NIE, 1978). Studies under way, conducted for the Office of Education, show that slightly more than 90 percent of the schools with high concentration of students with minority background receive some kind of compensatory aid, either Federal or State funds. Of the Nation's elementary schools 68 percent obtained Title I aid. An additional 14 percent received other funds, mostly State funds with the same purpose. Of eligible children in eligible schools 67 percent were reached. By and large, the Title I funds to a remarkable extent were targeted on schools with a high concentration of poor students. But the fact that a considerable number of educationally disadvantaged were not beneficiaries, as well as the

fact that quite a few who were well off were serviced, raises questions about the adequacy of the allocation criteria, a matter to which I shall return later.

Title VII of ESEA, the Bi-lingual Education program, deserves special mention. It was launched in 1967 at the initiative of legislators from California and Texas with the aim of improving education of children from non-English speaking homes, mostly Hispanics. The provisions were designed to support schools in providing bi-lingual instruction. Although in theory funds were targeted for children with limited proficiency in English, the Bi-lingual Education Act has popularly been identified as a Mexican-American project because of its emphasis on teaching children of Mexican (and Puerto Rican) background.

During our site visits, particularly in New Mexico, it was apparent that the term "disadvantaged" and even more the term "compensatory" were resented in connection with bi-lingual education. A child from a Spanish-speaking home in New Mexico who comes from a family that for centuries has lived there and who accordingly has a strong cultural identity is not necessarily "disadvantaged" in terms of lacking basic material and cultural assets. Nor can one say that it has to be "compensated" for not having English as a mother tongue.

There are among those who operate the bi-lingual programs, and to some extent among those who are beneficiaries, two opposing philosophies about the main objectives of the program. On the one side, there are those who argue that children from non-English speaking homes should become proficient in English as rapidly and efficiently as possible even at the price of losing contact with their home language and what that implies of alienation from their cultural heritage. On the other hand, there are those who regard bi-lingual education as a means of maintaining their cultural identity. A similar controversy has emerged in Western Europe where many millions of children of the "guest-workers" have forced the educational authorities to take steps in providing adequate education for them (Rist, 1978).

Considering the fact that the Hispanics soon will constitute the largest minority in the United States it would have been in order if the Bilingual Education Act had been included in the efforts to evaluate compensatory education.

While the OECD review team in the fall of 1978 conducted its site visits, the 95th Congress was in its final phase of discussing the amendments to the various titles in the ESEA, which was up for a 5-year reauthorization (Education Amendments, 1978). In the first place, the Amendments aimed at simplifying the

paper work which, as was repeatedly noticed during the site visits, was sometimes excessive, detailed, and burdensome to the local authorities. Instead of submitting applications annually it can now be done triannually. The Amendments of 1978 aim at better targeting than before on areas with high concentration of needy children. Thus, the basic Federal grants to the States are supplemented over and above the 1979 appropriations by extra funds in relation to the number of families in the State with an income below the National median according to the 1975 survey conducted by the Bureau of the Census. Furthermore, the basic allocation formula will, beginning in 1980, take 100 per cent of the children from families receiving AFDC instead of only 2/3 which has been the provision since 1965. Special grants are made available to school districts with a high concentration of children from low-income families. Thus, counties where the number of such children exceeds 20 per cent of the total are eligible for such grants.

Federal legislation on compensatory education has right from the outset been very strict on two principles: Federal funds should "supplement not supplant", and the States and local educational authorities should be able to demonstrate "maintenance of effort", that is to say, not slacken their efforts in terms of general funding of elementary and secondary education.

A meticulous control of the pursuance of these principles has resulted in excessive red tape. Exclusion from the first principle can now occur under the new provisions, in case State funds are used for compensatory education programs similar to Title I. The waiver provision for maintenance of effort has been modified to allow waiver from one fiscal year in case of "exceptional and unforeseen" circumstances. Thus, a disruption from one fiscal year to another of a useful program, simply because a district or a school happened to lose a fraction of a percent of eligible children and thereby falls below the level required to receive funds, is avoided.

Parental participation, a cornerstone in the original ESEA legislation, did not always work out according to anticipation. This might have been the reason why it was included in the NIE Compensatory Education Study (NIE, 1977f) which in turn was based on the 1975-76 NIE National Survey of Compensatory Education, which in its turn was based on efforts by the Stanford Research Institute (1971), Vanecko et al. (1977) of Abt Associates, and Goettel and Kaplan (1977) of the Syracuse Research Corporation. Parental Advisory Councils (PACs) have had considerable difficulties in finding their identity and have in many places become mere formalities. In other places they have become the focus of political patronage and struggle. Therefore, the regulations in the 1978 Education Amendments

require certain procedures for selection of parents on the PACs, a minimum number of meetings per year, and a minimum number of individuals to serve on the PACs.

Sometimes there has been an overzealous formalism with regard to which children in a given school or class may be served. There has been a tendency to categorize certain teachers as Title I teachers versus the "regular" ones. In order to reduce such tendencies, all children in a school where 75 per cent of them are eligible for compensatory education can be served. In a similar vein, teachers who are employed mainly for compensatory programs are required to spend a certain amount of their time on other activities which belong to their teaching role, such as consultations with other teachers and with parents and students.

Targeting of Compensatory Education.

The difficulties in implementing Federal commitments to the education of the disadvantaged can be fully appreciated only if one realizes the enormous complexities involved all the way from policy making, preparation of regulations, assessing needs for support, distribution of funds, planning programs at the local level, and evaluating and accounting upwards. Federal aid in education on a massive scale has a short history. The planning of legislation

and the launching of the ensuing programs occurred within a short time span in the early 1960's. As always, once certain practices have been established and entrenched, it is difficult to change them. Any more deepgoing change in the administrative machinery is felt as an intrusion upon established empires and spheres of influence. In addition, politicians take particular interest in programs that in a tangible way affect their particular constituencies. The OECD team could not avoid noticing how keenly aware congressmen were about the effects that proposed changes in the allocation criteria and regulations could have on their constituencies. This explains the reluctance to change existing practices of distributing funds.

Compensatory education funds under Title I of ESEA are distributed according to an allocation formula Congress agreed upon in the 1960's and which, with minor modifications, is still employed. The poverty statistics are obtained from census data that could be more than ten years old. This means that in certain areas the composition of the population might have changed rather drastically in a society as mobile as the American. Funds are first allocated to the State, which in its turn allocates them to the various school districts which then allocate them to the various school sites.

The number of poor, as assessed by census

data at the county level, is the first criterion of identification. Schools within districts are ranked according to incidence of poverty. Available funds are then distributed according to this rank order. This means that a considerable number of disadvantaged children, beyond those who live in areas or districts not eligible according to the poverty criterion, are not serviced. Thus, there are two categories of educationally disadvantaged children who are not covered by Title I provisions: (1) children in schools not eligible or not covered by the program, and (2) children in eligible schools for whom funds are insufficient. The NIE (1977c) study that addresses itself to this problem found that about one-third of the eligible children were not served by Title I programs.

Let me take one example from the site visits paid by the OECD team. When interviewing people at the Board of Education of the Los Angeles Unified School District we were told that by 1978 in L.A. only 172 out of more than 400 elementary schools and some 30 per cent of the eligible children were served by State programs. The criterion for distributing State aid to schools for compensatory education, which in California at least until then was comparatively high, was in L.A. a composite of (1) AFDC, (2) number of free school lunches, (3) property assessment, and (4) family income. After the eligible schools had been identified by poverty

criteria, students to be served were determined by the grade 6 reading scores over the last 4 years.

A survey on a national, representative sample conducted by the Systems Development Corporation (SDC), still under way when the OECD team visited Los Angeles, found that two-thirds of all elementary schools obtained Title I funds.

There is evidently a trade-off between horizontal coverage and concentration, which has a bearing on the practice of letting all children who score below the 50th percentile in eligible schools become eligible for services. We were informed that a study subcontracted to SDC shows that 50 percent of the children who are poor and score below the 50th percentile are serviced as compared to 35 percent who according to the formula are not poor and score in the same range. This gives reason to question the adequacy of poverty measures for identification of needy children.

Another aspect of mistargeting related to level of reading competence is the following. I learned at the interview session with SDC that the survey had shown that 12 per cent of the poor who score above the 50 percentile were serviced. The cutoff point was set at the 50th percentile, which seemed to be remarkably high

and was brought in question at the "confrontation meeting" in Paris. The 50th percentile as cutoff score is not consistent with the definition that the Office of Education employs in its 1977 Annual Evaluation Report to Congress: "The term (disadvantaged) applies to children and adults who typically cannot succeed in the traditional education systems and programs." (op.cit., p. 4). It would seem more in line with the intentions of the legislation to place the cutoff point for educational failures somewhere between the 15th and the 25th percentile. Such a practice would also be more compatible with the standards set in the minimum competency testing programs.

Unfortunately, the NIE evaluation studies have not addressed themselves to the problem of horizontal coverage and to what could be gained in services for the most needy with more concentration on them by a lower cutoff score. Again, the resistance from many school districts, who at least get some "sprinkling" of the funds, might have been very strong with repercussions on the respective congressmen.

The OECD group noted a striking lack of continuity in the vertical coverage of Title I programs and therefore in sustained efforts throughout the eligible students' entire school career. Compensatory education funds are concentrated at two completely separated stages of

the educational career. Title I funds are about 80 percent used in grades K through 3. In spite of the fact that there are no legislative barriers against using it at later stages, the funds under Title I provisions are almost entirely spent in the elementary school through grade 6. Very little is spent on compensatory education measures at the secondary level, where poor students from deprived background represent a serious problem. As long as students are in elementary school, attention is paid to the development of their basic skills in, for instance, reading. This objective seems to get out of focus at the junior high school level, and many students enter senior high school with very low reading competence, a fact which in many States has inspired recent legislation on minimum competency testing.

I have earlier pointed out the trade-off between horizontal coverage and concentration in distributing compensatory education funds. There is also a trade-off related to vertical coverage: the number of students that can be covered versus continuity and sustention of services. Looking back at the experiences education Ralph Tyler pointed out that "improvements in learning can be maintained and "improvements in learning can be maintained and increased when the program provides for a sequence of three to four years rather than one-shot efforts to help." (Tyler, 1974, p.

170). He estimated that bringing learning gains of the disadvantaged up to the national average would require a 50 percent higher per pupil expenditure than for the average student. Statements on the effect of increased sustention of efforts have been made by experts on early childhood education with particular reference to Head Start programs (Bronfenbrenner, 1974). It would seem that sustention could be bought at the price of reducing horizontal coverage by lowering the cutoff score.

During the 1970's five national commissions in the United States have dealt with problems of secondary education (Coleman et al., 1974; Brown et al., 1973; Martin et al., 1974; NASSP, 1972; Weinstock, ed., 1973). One official in HEW referred to the junior high school as the "disaster area in American education." It is the stage when neglect of basic skills begins to be strongly felt. At the end of this stage dropouts take place. But even before, the increased absenteeism means that dropout has de facto started.

Problems of dropout and unemployment among young people in the age range 16 through 21 are issues with serious implications for American education, as has been evidenced by the committee set up by the National Academy of Education and chaired by Clark Kerr (NAE, 1979). Bridges between the formal educational system

and the world of work have to be built in order to bring young people, who have grown up in a weak and unsupportive environment and who with Trow's typology are really deprived, into productive work. One notes that the Youth Employment Demonstration Act of 1977 conducts experiments in five cities with programs where young people are guaranteed either work or a place in school. It should also be mentioned that under the provisions of the Youth Act considerable funds are available for innovative projects linking school and work. Establishment of vocational-technical centers and the quest for building "life-role competencies" in senior secondary school are other examples of efforts in the same direction.

To achieve such objectives as better vertical coverage by sustained efforts and including upper elementary school as well as secondary school, a program integration which goes far beyond the confines of both the Office of Education and HEW has to be achieved. The establishment of a Department of Education has in this particular respect been a step in the right direction. But compensatory programs at the secondary level would require a better coordination between programs which until now have been operated in separation by HEW and the Department of Labor, respectively.

In the Instructional Dimensions Study NIE

(1977d) reports a 7-month follow-up of students in grades 1 and 3 in 14 school districts. Certain practices, such as "pullout instruction" as compared to "mainstream instruction," were studied. Since, as was noticed by the OECD team, pullout is overwhelmingly used, the results of such a comparison have central strategic importance. The study was also designed to assess the increments in student competence from the beginning of the fall through the end of the spring semester.

The study came out with two main findings. In the first place, it was found that compensatory education students, in the wording of the NIE Final Report, "make significant achievement gains over a school year period" (NIE, 1978, p. 79). These effects were larger than those obtained in previous studies of Title I students In interpreting this finding one should, however, keep two important conditions for the study in mind. Although the sample contained students from 400 classrooms in urban and rural districts it was by no means nationally representative and did not purport to be so. The Final Report points out that the "programs examined were purposely selected for certain instructional features and cannot be considered a representative sample of Title I reading and mathematics services." (NIE, 1978, p. 80). Secondly, and probably more important, 10 percent of the compensatory education students

whose initial test scores were above average, were excluded from the follow-up. As justification for doing, so it is pointed out that the rates of gain among the excluded students were "similar" to the ones among the rest. If compensatory education should make sense, those who score far below average ought to show considerable gains by being subjected to special treatment. We are now left without evidence about the effects that this special treatment leads to over and above those being achieved by regular instruction.

Evaluation of Compensatory Education Programs.

Needless to point out, evaluation exercises are by their very nature beset with value problems as well as with technical problems. It is difficult enough to identify and operationalize criteria supposed to measure the objectives one sets out to achieve in social intervention programs. But an even more troublesome problem, although of another kind, is encountered in the attempts to identify the objectives and the values embedded in these criteria. They are usually anchored in the middle of political controversy. Let me take one example from Sweden that relates to the legislation in the 1950's on a common, comprehensive school in Sweden (Husén, 1962). When an education bill proposing a change-over from a dual or parallel school system of basic education to a common, comprehen-

sive 9-year school was discussed by the Swedish Parliament in 1950, it met, as could be expected, resistance on the part of the conservatives. A Special Education Committee reviewed the Bill and presented its review to the Parliament, where unanimity was bought at the price of clarity. The first recommendation made by the Committee had the following wording: "Actions should be taken within a certain time, which is to be decided upon by the Parliament,...to introduce a comprehensive school based on nine years of mandatory school attendance. This school should, according to the extent the planned experiments prove its suitability, replace elementary school, the continuation school", etc. In order to achieve consensus the Committee Chairman had inserted the words "according to the extent...". The proponents of structural change thought that a decision had been taken to introduce the comprehensive school and that certain modifications could be envisaged depending upon the outcomes of the 10-year pilot program which was included in the Bill. The conservatives interpreted the recommendation to the effect that adoption of the very idea of the comprehensive school should depend on whether the pilot program proved to be "superior" in terms of making its students more "competent" than the previous types of schools, particularly the selective academic lower secondary school. Since a decision was taken that the pilot program should be accompanied by ex-

tensive comparisons between the new and the traditional types of schools, the ensuing evaluation surveys during the 1950's were followed with keen interest by both opponents and proponents. The comparisons were made in terms of student cognitive competence in certain key subjects, and nobody questioned the validity of comparisons between school types within certain respects quite different objectives. The Social Democrats wanted a school which contributed to equalization and conceived the comprehensive school reform in social terms, whereas the conservatives discussed the reform in terms of instructional and pedagogical efficiency.

Similar problems were encountered in evaluating the Title I programs. I have earlier discussed the philosophy behind them. Extra funds to support disadvantaged children are expected to improve their educational attainments. This would in turn improve their coping power as adults, make them better prepared for working life, and lift them out of poverty. Therefore, the ultimate criterion of how "effective" Title I is would be adult occupational status and/or earnings of students serviced by the programs. This would require a longitudinal study which would administratively be difficult and costly to conduct - not to mention that it would not be attractive to politicians who prefer instant results within the short perspective. The length of their term of office tends to determine the

length of their perspective.

Therefore, practically all evaluation efforts have opted for short-range criteria, thereby using the one that is close at hand: student achievement. In most cases one has tried to assess what happens to the children in the program over one school year. Even that can be a tricky exercise, considering the high turnover of children in the categories serviced by Title I. When the turnover is about one hundred percent, which occurs in some schools, the quest for "sustained efforts" indeed tends to become an empty phrase.

The research literature on evaluation of various compensatory education programs, particularly Head Start, Follow Through, and Title I, is, indeed, extensive. If one includes the annual reports submitted by the local educational authorities and by the U.S. Commissioner of Education, the documentation is enormous. In addition, commissioned studies, many of which were designed in grand format and at great expense, have been conducted. It would be highly presumptuous to pretend to cover even the major relevant publications. Nor could I possibly pretend to come up with my own evaluation of studies into which impressive competence and considerable sums of money have been invested. I shall therefore mainly confine myself to general observations and to raising some issues.

There have been many small-scale evaluations of Head Start programs. The first more extensive and significant attempt to assess the impact of these programs was the Westinghouse Report of 1969. Little and Smith (1971), who were commissioned by OECD to conduct a study of the strategies of compensation employed in the United States, conclude from the Westinghouse Report: "The overall findings of the study suggested that the summer programs had little sustained effect; in general no differences could be distinguished between Head Start and non-Head Start children." (p. 52). Although Head Start children were slightly ahead of non-Head Start children at the beginning of grade one, by grade two there were no significant differences in school achievement. Similar results were obtained with regard to affective outcomes. As usual in studies of this kind, it was beset by methodological flaws, such as lack of pretest scores and one-sided emphasis on achievements. Tremendous variations between programs defied attempts to relate specific treatments to specific outcomes. Thus, so far the Head Start evaluations have not yielded any consistent and/or conclusive results that could serve as guidelines for policy makers. One should remember, however, that in this respect attempts to evaluate effects of treatments similar to the ones that go on in the regular school setting have also given inconclusive results, leading to the desperate conclusion that "schooling does

not seem to make any difference."

A massive attempt using a longitudinal research strategy has recently been launched by the Education Commission of the States. Preliminary findings seem to indicate more positive effects of early childhood intervention than found in previous investigations.

The impressive resources put at the disposal of the researchers commissioned to conduct evaluation studies have not been able to offset the shortcomings that stem from being put into the straight-jacket of policy makers and planners. The demand for quick and "relevant" results has tended to sterilize much of this research. The evaluation of the "planned variation" field experiments conducted under the auspices of the Follow Through and criticized by House et al. (1978) serves as a good illustration and shall therefore be dealt with somewhat more in detail.

The Follow Through program was launched as an attempt to sustain the effects achieved by Head Start in providing continuing help to former Head Start children as they entered regular school. Follow Through was originally conceived as a program providing additional compensatory education in Kindergarten and in the first three grades of elementary school for children who had previously participated in Head Start. In 1968

it was decided to make the program more experimental and to try out various models in early childhood education that should subsequently be evaluated (cf. Rivlin and Timpane, 1975). It was decided to try out about a dozen models, and 15 million dollars were allocated to the entire program. By 1978 the annual allocation had increased to 59 million. The total program costs over the decade since the inception amount to roughly 500 million dollars. Over the entire period 20 sponsors had worked with projects on 180 school sites. The Abt Associates, Inc. was commissioned to conduct the evaluation that comprised a sample of some 20,000 students who were followed up over a 4-year period. The total evaluation costs have been estimated to be somewhere between 30 and 50 million dollars. The evaluation exercise reported in half a dozen massive volumes has been subjected to criticism by House et al. (1978). Their report, sponsored by the Ford Foundation, as well as rebuttals from the researchers responsible for the Abt Associates evaluation have been published in a special issue (No. 2, 1978) of the Harvard Educational Review. This issue makes an extremely interesting reading, because it amply illustrates the point above on the relationship between educational research and educational decision-making. Researchers were in the first place called upon to make their contribution ex post facto, when the models had been decided upon and with little possibility of influencing

either the design of the "planned experiments" or the way they were conducted. Very often the conduct of the experiments was arbitrarily changed due to particular local circumstances and interests that had little to do with the aims of the "experiments".

A host of troublesome methodological problems were encountered in the attempts to evaluate conclusively the Follow Through programs. A few major problems and flaws deserve to be mentioned. In doing so I am largely drawing upon the report by House et al. (1978). There was no systematic selection of school sites and/or teachers within schools. The control groups were grossly mismatched with the experimental groups. Adequate evaluation instruments were not developed from the outset and focused mainly on cognitive competence. On the whole the outcome domains were poorly covered. The goals implicit in the various experiments within the same models were so diverse that it was very difficult, not to say impossible, to develop common instruments that could adequately cover the entire range. The aggregation level represented another matter of concern: the student instead of the classroom was used as the unit of analysis. This was inconsistent with the statistical technique used in adjusting the means of the control groups by means of analysis of covariance. Half the original sample of children was lost during the 4-year follow up. The

hopeless situation for the evaluators was reflected in the fact that the variation in mean achievement within models was greater than between models. This reflects the role played by contextual factors for the outcomes. Such conclusion is drawn by the Abt Associates, Inc. researchers (Anderson et al., 1978) in their reply to the criticism leveled by House et al. (1978):

> "None of the seventeen models in the evaluation demonstrated that it could compensate consistently for the academic consequences of poverty... (We) conclude that the Follow Through strategy of externally sponsored curricular change is not a reliable tool for raising the test scores of poor children... Local circumstances and behavior clearly have more to do with children's test performance than do the intentions, theories, and rhetoric of outside interveners.... Taken as a whole, the evidence presented in our report (Stebbins et al., 1977) suggests that if a local school system has the potential for effective compensatory education, then outside resources of the Follow Through kind can sometimes catalyze this potential. But if not, intervention seems likely to be somewhat disruptive and counterproductive." (Anderson et al., 1978,

p. 162).

This statement is in line not only with common sense but also with findings from research on the determinants of educational innovations. The contextual factors, the ecology of the educative process going on in the school, have to be taken into account. The involvement of the parents and the extent to which they in cooperation with the school are participating in the education of their children is of pivotal importance.

Congress in its 1974 Education Amendments requested, as mentioned earlier, the National Institute of Education to conduct a study of compensatory education, its "fundamental purposes and effectiveness." In time for consideration in connection with the reauthorization of the compensatory education programs a final comprehensive report was submitted in September 1978 (NIE, 1978). The NIE reports reflect a shift towards a widened perspective in conducting evaluation. The NIE compensatory education study is in the first place not an "in-house" one in terms of original research endeavor with what that implies of research design, data collection, and statistical analyses, but rather an attempt to collate the extensive research conducted in some 35 projects, most of them commissioned by the Federal Government and conducted by outside organizations.

As has been repeatedly pointed out, most of the evaluation projects have mainly been preoccupied with the effects of the compensatory education programs on student cognitive development as measured by achievement test scores in the basic skills. This has to be seen in the context of the emphasis on such skills not only in the legislation but also in the actual operation of Title I programs.

The bulk of the Title I resources has gone into remedial reading. Recent "back-to-the-basics" battle cries have reinforced the tendency to look at the outcomes in terms of student performance, particularly in reading and mathematics. Thus, the effective and social development of the students, which, of course, is less tangible and therefore more difficult to measure, has tended to be neglected. Not least the Follow Through evaluation exercise points at the importance of changes in the "ecology" or "climate" in schools and communities where compensatory programs have operated. Parental involvement in the education of the children was enhanced by the Follow Through, where parents twice as often as in the control group volunteered to help in the school. Many low-status parents became interested in improving their own education. Another important by-product seems to have been that some workable comprehensible models of primary education were developed under the auspices of Follow Through.

The program could thereby serve as a kind of curriculum and instrument development laboratory. A third by-product seems to have been the opportunity that various planned variations experiments have provided for effective on-the-job inservice training for teachers.

During a visit to Systems Development Corporation the OECD team had the opportunity to meet with the researchers commissioned to conduct a comprehensive study of the sustained effects of Title I basic skills. Base line data collected for a nationally representative sample of students and parents of a considerable size make it possible to survey on the basis of what criteria services in the Title I programs are delivered. This is an impressive attempt to overcome some of the technical shortcomings that have beset some previous studies. It also tries to map out what goes on in the home between parents and children which, regrettably, because of the size of the sample, has to be studied rather extensively. In our conversation with the researchers we learned that the longitudinal emphasis, which was a major purpose for launching the study, might because of the high costs have to be played down. One must regret this if it occurs and hope that it might be possible to conduct a follow-up study on at least a subsample in order to assess the sustained effects of the compensatory education programs.

Concluding observations.

Both in reviewing policies for compensatory education and in the attempts to evaluate the outcomes a certain amount of generosity -- and humility -- on the part of the researcher has to be exercised. Legislation and its implementation in terms of regulations are outcomes of political compromise, a process where representatives of States and districts have tried to maximize the benefits accruing to their own constituencies. One cannot, therefore, in the legislative wording expect the same clarity and consistency of definition as in a graduate seminar of, for instance, what is meant by "equality" or "disadvantage" and how these terms should be operationalized when it comes to social intervention.

In looking back on the development of American education over the last two decades one is deeply impressed by the considerable strides that have been made, not least by Federal initiatives, which have contributed to putting vital issues, particularly the education of the disadvantaged, on the national agenda. In spite of frustrated expectations, setbacks of programs, criticism of falling standards, and concerns about the young people who leave school as functional illiterates, the faith in education among the American people has remained remarkably unaffected. As an observer from abroad,

one cannot avoid noticing that the Proposition 13 movement is not directed against education as such but against big government (as has also been evidenced by opinion polls). One indication of the sustained faith is the fact that the 1978 Federal appropriations for education went up 16 percent, which was far beyond what was needed to make up for inflation.

With regard to evaluation I would like to make two observations. In the first place, a broader perspective has to be employed in looking at the outcomes of intervention programs than in some of the commissioned studies. By and large the employment of a wider range of criteria of evaluation is called for. Irrespective of whether one accepts or rejects the traditional experimental model as feasible and adequate in assessing the outcomes of planned intervention in the educational and social welfare demand, the problem of keeping the relevant factors "under control" is, indeed, formidable and is among other things reflected in the enormous costs incurred in evaluating the Follow Through models. I agree with the conclusion made by the group of researchers commissioned to evaluate the Follow Through program in saying: "Poor children still tend to perform poorly in school even after the best and the brightest theorists -- with the help of parents, local educators, and the federal funds....have done their best to change the situation." (Anderson et al., 1978,

p. 163). Secondly, even taking into account the difficulties of conducting longitudinal studies because of the high turnover of students and of keeping track of people who move, I think that wellplanned, small-scale longitudinal studies conducted as case studies where different models of compensatory education are operated, would prove more than large-scale surveys.

We are, in spite of all efforts by educational research, just at the beginning of our attempts to assess what the school is doing to its students. Until now, our methods have been extremely crude and limited to easily measured cognitive performances. Our notions of what makes a difference in student competence have also been rather crude because of our failure to assess what, for lack of a better expression, could be called the "climate" of the school or the classroom. The study of a group of inner London schools conducted by Michael Rutter (1978) marks a turning point in this respect. The rather meager, not to say, discouraging results from the evaluation exercises should not give rise to pessimism. In his final remarks at the Paris "confrontation meeting" the Chairman of the OECD team Peter Karmel pointed out that we lack methods of assessing the "chemistry" or "climate" of the school. Furthermore, we tend to overlook that increments in educational outcomes usually are very small. We cannot expect minor changes in input to produce

dramatic results. The compensatory education programs constitute after all small changes in the larger context of formal education.

Finally, it should be recognized that the educational system cannot bear the whole burden of bringing about social change. Educational reforms cannot have an impact and cannot successfully be implemented if they are conceived as if they were operating in a social vacuum. In order to succeed they have to be parts of overarching social reforms, for which they cannot substitute.

REFERENCES

Anderson, Richard B., Robert B. St. Pierre, Elizabeth C. Proper and Linda B. Stebbins (1978) "Pardon Us, But What Was That Question Again?: A Response to the Critique of the Follow Through Evaluation". Harvard Educational Review, Vol. 48, No. 2, 161-170.

Bronfenbrenner, Urie (1974) "Is Early Intervention Effective?". In Hope Jensen Leichter (ed.), The Family as Educator. New York: Teachers College Press.

Brown, B. F. et al. (1973) The Reform of Secondary Education. A Report of the National Commission on the Reform of Secondary Education. New York: McGraw-Hill.

Coleman, James S. et al. (1974) Youth: Transition to Adulthood. Report of the Panel on Youth of the President's Science Advisory Committee. Chicago: University of Chicago Press.

Education Amendments of 1978: Conference Report. 95th Congress 2nd Session. Report No. 95-1753. House of Representatives, 10th October, 1978.

Goettel, Robert J. and Bernard A. Kaplan (1977) A Study of the Administration of the Elementary and Secondary Education Act (ESEA) Title I in Eight States. Syracuse, New York: Syracuse Research Corporation.

Hechinger, Fred M. (1976) "Public Education as the Great Equaliser". In Nelson F. Ashline et al. (eds.) Education, Inequality and National Policy. Lexington, Mass.: Lexington Books.

House, Ernest R., Gene V. Glass, Leslie D. McLean, and Decker F. Walker (1978) "No Simple Answer: Critique of the Follow Through Evaluation." Harvard Educational Review, Vol. 48, No. 2, 128-160.

Husén, Torsten (1962) Problems of Differentiation in Swedish Compulsory Schooling. Stockholm: Svenska BokforlagetScandinavian University Books.

Husén, Torsten (1975) Social Influence on Educational Attainment. Paris: OECD/CERI.

Husén, Torsten (1978) "Educational Research and Educational Reform: A Case Study of Sweden." In Patrick Suppes (ed.) Impact of Research on Education. Washington, D.C.: National Academy of Education, 523-579.

Husén, Torsten (1979) The School in Question. London and New York: Oxford University Press.

Jencks, Christopher et al. (1972) Inequality: A Reassessment of the Effect of Family and Schooling in America. New York and London: Basic Books.

Jencks, Christopher et al. (1979) Who Gets Ahead? The Determinants of Economic Success in America. New York: Basic Books.

Keppel, Francis (1966) The Necessary Revolution in American Education. New York and London: Harper & Row.

Little, Alan and George Smith (1971) Strategies of Compensation: A Review of Educational Projects for the Disadvantaged in the United States. Paris: OECD/CERI.

Mann, Ada Jo, Adele Harrell, and Maure Hurt, Jr. (1977) A Review of Head Start Research Since 1969 and an Annotated Bibliography. Washington, D.C.: HEW Publication No. (OHDS) 78-31102.

Martin, John H. et al. (1974) Report of the National Panel on High Schools and Adolescent Education. Washington, D.C.: U.S. Office of Education.

NAE (1979) Education for Employment: Knowledge for Action. By a Task Force on Education and Employment; Clark Kerr, Chairman. Washington, D.C.: Acropolis Books.

NASSP (1972) American Youth in the Mid-Seventies. Washington, D.C.: National Association of Secondary School Principals.

NIE (1977a) Administration of Compensatory Education. Washington, D.C.: National Institute of Education, U.S. Department of Health, Education, and Welfare.

NIE (1977b) Compensatory Education Services. Washington, D.C.: National Institute of Education, U.S. Department of Health, Education, and Welfare.

NIE (1977c) Demonstration Studies of Funds Allocation Within Districts. Washington, D.C.: National Institute of Education, U.S. Department of Health, Education, and Welfare.

NIE (1977d) The Effects of Services on Student Development. Washington, D.C.: National Institute of Education, U.S. Department of Health, Education, and Welfare.

NIE (1977e) Title I Funds Allocation: The Current Formula. Washington, D.C.: National Institute of Education, U.S. Department of Health, Education, and Welfare.

NIE (1977f) Using Achievement Test Scores to Allocate Title I Funds. Washington, D.C.: National Institute of Education, U.S. Department of Health, Education, and Welfare.

NIE (1978) Compensatory Education Study: A Final Report from the National Institute of Education. Washington, D.C.: The National Institute of Education, U.S. Department of Health, Education, and Welfare.

OE (1979) U.S. Federal Policies for the Education of the Disadvantaged. Washington, D.C.: Office of Education, U.S. Department of Health, Education, and Welfare.

OECD (1971) Reviews of National Policies for Education: France. Paris: OECD.

OECD (1972) Reviews of National Policies for Education: Germany. Paris: OECD.

OECD (1980) Reviews of National Policies for Education: United States. Paris: OECD.

Report of the President's Task Force on Education (1964). Washington, D.C.: The White

House, 14th November, 1964.

Rist, Ray (1978) Guestworkers in Germany: The Prospects for Pluralism. New York: Praeger.

Rivlin, Alice M. and P. Michael Timpane (eds.) (1975) Planned Variation in Education: Should We Give Up or Try Harder? Washington, D.C.: The Brookings Institution.

Rutter, Michael (1979) Fifteen Thousand Hours. London: Open Books.

Sewell, William H., Robert M. Hauser, and David L. Featherman (eds.)(1976) Schooling and Achievement in American Society. New York: Academic Press.

Stanford Research Institute (1971) Implementation of Planned Variation in Head Start: Preliminary Evaluations of Planned Variation in Head Start According to Follow Through Approaches (1969-1970). Stanford Research Institute, May 1971.

Trow, Martin (1978) Reflections on Youth Problems and Policies in the United States. University of California, Berkeley: Center for Studies in Higher Education.

Tyler, Ralph W. (1974) "The Federal Role in Education." The Public Interest, No. 34, Winter 1974, 164-187.

Vanecko, J.J., F.X. Archambault, and N.L. Ames (1977) ESEA Title I Allocation Policy: Demonstration Study. Cambridge, Mass.: Abt Associates, Inc.

Weinstock, Ruth (ed.)(1973) The Greening of the High School. New York: Educational Facilities Laboratory.

Westinghouse Learning Corporation (1969) The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development; Executive Summary. Ohio University: Westinghouse Learning Corporation.